

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 15:09:08 : Search time 968.083 Seconds
(without alignments) 6775.784 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486
Sequence: 1 atgacacagagatgttgcga.....aagtcattcaacagaaaa 486

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estinu:*
5: em_estcov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53.8	11.1	998	12	BH133989 ENTNM17TF
2	52.4	10.8	868	12	BH162606 ENTNV08TF
3	49.6	10.2	279	12	CNS03W9F
4	47.2	9.7	845	12	AZ685166 ENTLG11TF
5	46.4	9.5	325	12	CNS03PK8
6	46.4	9.5	624	10	BT001798
7	46.4	9.5	872	10	BT009355
8	46	9.5	909	12	CNS00UTL
9	46	9.5	939	12	CNS00CNG
10	44.6	9.2	890	12	BH164851
11	43.2	8.9	427	9	AU052962
12	43.2	8.9	440	9	AU052941
13	42.6	8.8	442	10	BE801262
14	42.6	8.8	452	10	BT471035
15	42.6	8.8	536	12	CNS0300K
16	42.2	8.7	748	10	BT176637
17	42.2	8.7	805	12	AZ753057

C 18	41.8	8.6	723	12	AZ627893	AZ627893 IM0469118
C 19	41.8	8.6	745	12	CNS01MHE	AL150931 Anopheles
C 20	41.6	8.6	294	12	AU060898	AU060898 AU060898
C 21	41.2	8.5	1101	12	CNS00LOO	AU068607 Drosophila
C 22	41.1	8.4	328	9	A1437626	A1437626 sa37e06.y
C 23	41.1	8.4	344	10	BC352395	BC352395 sac03d01
C 24	41.1	8.4	352	9	A1443457	A1443457 sa33h07.y
C 25	41.1	8.4	403	9	A1442486	A1442486 sa27f06.y
C 26	41.1	8.4	431	10	BE555999	BE555999 sp96d11.y
C 27	41.1	8.4	496	10	BE805704	BE805704 ss48f09.y
C 28	41.1	8.4	509	9	AW734942	AW734942 sK93b01.y
C 29	41.1	8.4	520	10	BM093415	BM093415 sa108901
C 30	41.1	8.4	521	10	BE790034	BE790034 sa62c03
C 31	41.1	8.4	536	10	BT784659	BT784659 sa57f12
C 32	41.1	8.4	543	10	BT701733	BT701733 sa119e04
C 33	40.8	8.4	805	12	CNS04FW2	AL030427 Tetradon
C 34	40.6	8.4	596	12	BH483310	BH483310 BOGTE35TF
C 35	40.6	8.4	676	10	BE603900	BE603900 GS104 GS
C 36	40.6	8.4	861	12	CNS0075A	AL066834 Drosophila
C 37	40.4	8.3	173	10	BM159224	BM159224 EST561747
C 38	40.4	8.3	250	9	AU060486	AU060486 AU060486
C 39	40.4	8.3	444	9	AU033804	AU033804 AU033804
C 40	40.4	8.3	559	12	AZ304119	AZ304119 IM0004M04
C 41	40.4	8.3	576	12	AZ379826	AZ379826 IM0135E14
C 42	40.4	8.3	650	12	BH393056	BH393056 AG-ND-103
C 43	40.2	8.3	1101	12	CNS00LT2	AL078714 Drosophila
C 44	40	8.2	768	12	A0690676	A0690676 nbx00082B
C 45	40	8.2	1026	12	CNS003KS	AL064935 Drosophila

ALIGNMENTS

RESULT 1
LOCUS BH133989 998 bp DNA linear GSS 07-AUG-2001
DEFINITION ENTNM17TF Entamoeba histolytica sheared DNA Entamoeba histolytica genomic DNA sequence.
ACCESSION BH133989
VERSION BH133989.1 GI:15093050
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 998)
Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
Unpublished (2001)
JOURNAL
COMMENT
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 17
High quality sequence stop: 490.
Location/Qualifiers
1..998
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica sheared DNA"
/note="Vector: pHost; Site: 1. Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

BASE COUNT	ORIGIN	106 a	155 c	2 g	13 t	3 others
Query Match	10.2%	Score 49.6:	DB 12:	Length 279:		
Best Local Similarity	56.8%	Pred. NO. 0.0052:				
Matches	88:	Conservative	1:	Mismatches	66:	Indels 0: Gaps 0:
BASE COUNT	ORIGIN	106 a	155 c	2 g	13 t	3 others
Query Match	10.2%	Score 49.6:	DB 12:	Length 279:		
Best Local Similarity	56.8%	Pred. NO. 0.0052:				
Matches	88:	Conservative	1:	Mismatches	66:	Indels 0: Gaps 0:
BASE COUNT	ORIGIN	106 a	155 c	2 g	13 t	3 others

Query Match	9.7%	Score 47.2;	DB 12;	Length 845;
Best Local Similarity	52.6%;	Pred. No. 0.032;	Mismatches 93;	Indels 0;
Matches 103;	Conservative 0;	Mismatches 93;	Indels 0;	Gaps 0;
OY	74	TTATGCTGTCAAGAAATCGAGACTGGACCTGGGGCGCTTGTGTTCGCCGTGTGCACCAGC	133	
Db	743	TTGGTCCTGTTGTGTATCGTAGTGCTGTATTGTCGGTGTGATGAGTGCCTGTGGTGTGCTG	684	
OY	134	TGTTTTGTAAAGATAGTGTGTCGATTCGAACAGCCATTGTGTCGCTGTTTTGGTGCAT	193	
Db	683	TTGGTGTGCTGCTTTGTGTGTGATGTTGGTGTCGTCGTTGTGTGTGCTCATTTGCTGTG	624	
OY	194	TAAATGTTCTTAATAATCGTCAATCGATGATGCACAGAGATTAATAAAGCAAAACACA	253	
Db	632	TTGTTGGTGTGCTTGTGTGTGCTGTACTGATGTATTATAAAAAGCAAGTAACACAAAAA	564	
OY	254	GTTTGCAAAAAGGTAA 269		
Db	563	TTGAAAAATTAAATAA 548		
RESULT 5				
CNS03PK8/c		325 bp	DNA	linear
LOCUS				GSS 17-MAY-2000
DEFINITION				Tetraodon nigroviridis genome survey sequence PUC-ori end of clone 045118 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION				AL254753.1 GI:7975765
VERSION				GSS; genome survey sequence.
KEYWORDS				Tetraodon nigroviridis.
SOURCE				Tetraodon nigroviridis.
ORGANISM				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neteleostei; Acanthomorphia; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE				1 (bases 1 to 325)
AUTHORS				Roeest-Crollius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE				Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL				Unpublished
REFERENCE				2 (bases 1 to 325)
AUTHORS				Roeest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Winkler,P., Brotlier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE				Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL				Unpublished
REFERENCE				3 (bases 1 to 325)
AUTHORS				Genoscope.
TITLE				Direct Submission
JOURNAL				Submitted (12-Apr-2000) to the EMBL/GenBank/DBJ databases
COMMENT				This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES				Location/Qualifiers
source				1..325
				/organism="Tetraodon nigroviridis"
				/db_xref="taxon:99883"
				/clone="045118"
				/clone_1lb="G"
				/note="Genoscope sequence ID : C0BG045BD09SP1-end ; PUC-ori"
BASE COUNT				157 a 157 c 10 g 0 t 1 others
ORIGIN				
Query Match	9.5%	Score 46.4;	DB 12;	Length 325;
Best Local Similarity	52.6%;	Pred. No. 0.04;	Mismatches 91;	Indels 0;
Matches 101;	Conservative 0;	Mismatches 91;	Indels 0;	Gaps 0;

[illegible]

Query Match	8.8%	Score 42.6	DB 10	Length 442
Best Local	Similarly	53.3%		
Matches	90	Conservative	0	Mismatches 79
				Indels 0
				Gaps 0

```

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1050-3707"
/clone.lib="Gm-cl050"
/tissue.type="leaf tissue at various developmental stages
of greenhouse grown plants"
/dev_stage="3 week old"
/lab_host="DH10B"
/notice="vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
XhoI; The Clark N1L was constructed and seed was provided
by Dr. J. Specht, University of Nebraska (Shoemaker and
Specht, 1995). The cDNA library was constructed from mRNA
isolated from leaf tissue at various developmental stages
of 3 week old greenhouse grown plants. Complementary DNA
was synthesized from mRNA using a primer consisting of a
poly(dT) sequence with a XhoI restriction site and a 3'
anchor. EcoRI adaptors were ligated to the blunt-ended
cDNA fragments followed by XhoI digestion. The cDNA
fragments were directionally cloned into the EcoRI-XhoI
restriction site of the pBluescript vector. The ligated
cDNA fragments were transformed into DH10B host cells
(Glodbok). The library was constructed in cooperation
with Dr. Paul Keim's laboratory at Northern Arizona

```

BASE COUNT University."

ORIGIN 111 a 193 c 67 g 81 t

Query Match

Best Local Similarity 8.8%; Score 42.6; DB 10; Length 452;
Matches 90; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 35 TTATATCATGCTGTTTATGTCGTCGCCAAGACTTTAGTCGCAAGAGTCGAG 94
 DB 304 TTGCTGTGGCGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 94
 QY 95 CTGCGACTGGGCGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 245
 DB 244 CTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 154
 QY 155 GAGTTGCAGATGCCATTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 185
 DB 184 ATCTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 203

RESULT 15

CNS0300K/c

LOCUS

DEFINITION

Tetraodon nigroviridis genome survey sequence 17 end of clone

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

DB 423 GTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 111
 QY 137 TTGCTAAGTATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 364
 DB 363 GTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 111
 QY 197 TTGCTTCAAAATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 213
 DB 303 TTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 287

Search completed: October 27, 2002, 18:34:51
 Job time : 973.083 secs

BASE COUNT 163 a 236 c 79 g 43 t 15 others
 ORIGIN

Query Match

Best Local Similarity 8.8%; Score 42.6; DB 12; Length 536;
 Matches 78; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 77 GTCTCAGAGATGCGACTGCGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 136

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 16:06:08 : Search time 35.5337 Seconds
(without alignments)
3359.570 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486
Sequence: 1 atgaacagagggatgttgc.....aagtcattcacaagaaaa 486

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued_Patents_NA:*

- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	47.8	9.8	7218	1	US-08-232-463-14 Sequence 14, Appl
2	42.6	8.8	465	4	US-08-476-102A-2 Sequence 2, Appl
3	35.8	7.4	303	4	US-08-556-978B-80 Sequence 80, Appl
4	35.8	7.4	303	4	US-08-556-978B-81 Sequence 81, Appl
5	35.4	7.3	303	4	US-08-556-978B-82 Sequence 82, Appl
6	35.2	7.2	198	5	PCT-US95-10668-3 Sequence 3, Appl
7	35.2	7.2	198	5	PCT-US95-10668-4 Sequence 4, Appl
8	34.4	7.1	1046	1	US-08-361-467B-4 Sequence 4, Appl
9	34.4	7.1	1046	1	US-08-484-332C-4 Sequence 4, Appl
10	33.6	6.9	198	5	PCT-US95-10668-1 Sequence 1, Appl
11	33.6	6.9	198	5	PCT-US95-10668-2 Sequence 2, Appl
12	33.2	6.8	1011	2	US-08-825-781-2 Sequence 2, Appl
13	33.2	6.8	2072	2	US-09-073-362-2 Sequence 2, Appl
14	33.2	6.8	2072	2	US-09-243-920-2 Sequence 2, Appl
15	33.2	6.8	377	2	US-08-332-765A-1 Sequence 1, Appl
16	32.8	6.7	1690	2	US-08-276-452A-24 Sequence 24, Appl
17	32.8	6.7	1690	2	US-08-798-744-24 Sequence 24, Appl
18	32.4	6.7	1984	1	US-07-885-970A-25 Sequence 25, Appl
19	32.4	6.7	1985	1	US-08-298-687A-25 Sequence 25, Appl
20	32.4	6.7	1985	1	US-08-298-829-25 Sequence 25, Appl
21	32.2	6.6	1561	1	US-07-968-971A-3 Sequence 3, Appl
22	32.2	6.6	1561	1	US-07-824-247-44 Sequence 44, Appl
23	32.2	6.6	1561	1	US-08-142-473A-4 Sequence 4, Appl
24	32.2	6.6	1561	1	US-08-469-203A-4 Sequence 4, Appl
25	32.2	6.6	1561	1	US-08-469-203A-4 Sequence 4, Appl
26	32.2	6.6	1561	3	US-08-470-204A-44 Sequence 44, Appl
27	31.8	6.5	379	1	US-08-145-617-5 Sequence 5, Appl

28	31.8	6.5	2793	1	US-08-209-747-1 Sequence 1, Appl
29	31.8	6.5	2793	1	US-08-458-298-1 Sequence 1, Appl
30	31.6	6.5	5496	4	US-08-462-284-1 Sequence 1, Appl
31	31.6	6.5	1632	1	US-08-324-243-34 Sequence 34, Appl
32	31.6	6.5	1632	1	US-08-532-390-34 Sequence 34, Appl
33	31.6	6.5	1632	1	US-08-717-294-34 Sequence 34, Appl
34	31.6	6.5	1632	5	PCT-US95-11511-34 Sequence 34, Appl
35	31.4	6.5	925	3	US-08-858-003-1 Sequence 1, Appl
36	31.4	6.5	925	3	US-09-078-166-1 Sequence 1, Appl
37	31.4	6.5	925	4	US-08-997-467-1 Sequence 1, Appl
38	31.4	6.5	3254	4	US-08-965-903B-1 Sequence 1, Appl
39	31.4	6.5	3381	4	US-09-336-447A-6 Sequence 6, Appl
40	31.2	6.4	289	4	US-09-007-005-17 Sequence 17, Appl
41	31.2	6.4	289	4	US-09-244-796-17 Sequence 17, Appl
42	31.2	6.4	744	4	US-08-936-165A-75 Sequence 75, Appl
43	31.2	6.4	1086	1	US-08-415-751-47 Sequence 4, Appl
44	31.2	6.4	2242	1	US-08-641-627A-37 Sequence 37, Appl
45	31	6.4	1897	1	US-08-453-477-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
TITLE OF INVENTION: FALKNER, F. G.
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 1800 diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
CLASSIFICATION: 435
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pC-F1s
US-08-232-463-14
Query Match 9.8% Score 47.8: DB 1: Length 7218:

Best Local Similarity 2.0%; Pred. No. 2.5e-05;
Matches 7; Conservative 206; Mismatches 138; Indels 0;

Db 136 TTGTGTAAGTAGTGGTCAATGTCATGCGCATTTGGTGGTGGTGA 195
1442 TTGTGTAAGTAGTGGTCAATGTCATGCGCATTTGGTGGTGGTGA 195
OY 196 ATGCTGCTAAATCGGTCAATGTCATGCGCATTTGGTGGTGGTGA 195
Db 1382 RRR 1383
OY 256 TTGGAAGGTAAAGCAAGCAAGTGCACAGTGGCGTATCCAGATAGT 315
1322 RRR 1263
OY 316 TATAGTGTGAGCGCTGCTTACCAAGTTCATACCAAGCAAGTGC 375
Db 1262 RR 1203
OY 376 TATGCTGAGATTTGCAAAAGGCGATGTCAGGCAAGCAAGATTTAC 435
Db 1202 RR 1143
OY 436 ACTGATGCGGACCGGATGCTGTCGCAAGTTCATACCAAGCAAG 486
Db 1142 RR 1092

RESULT 2
US-08-476-102A-2
Sequence 2, Application US/08476102A
Patent No. 6355450
GENERAL INFORMATION:
APPLICANT: Fleischmann, et al.
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus
Patent No. 6355450
Influenzae Rd Genome, Fragments Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Ave.
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,102A
FILING DATE: 07-Jun-1995
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Marks, Michelle S.
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEX: 301-309-8439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 465 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Query Match
Best Local Similarity 54.9%; Score 42.6; DB 4; Length 465;
Pred. No. 0.00027;

Matches 84; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

OY 75 TACTCCTAGAAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 134
Db 171 TATCAAGGTGAGTGGTACGCTGGTGGAGTGGAGTGGAGTGGAGT 134
OY 135 GTTGTGAAGTGGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 230
Db 231 AATGGCGGTGGTGGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 194
OY 195 AATGGTCTGAATGCGTCAATGATGATCA 227
Db 291 AGCTGAGTAATAATGAGAAATAATGATCA 323

RESULT 3
US-08-556-978B-80
Sequence 80, Application US/08556978B
Patent No. 6268169
GENERAL INFORMATION:
APPLICANT: FAHNESTOCK, STEPHEN F.
TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898

COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/556,978B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/077,600
FILING DATE: JUNE 15, 1993
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AMMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9389-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-556-978B-80

Query Match
Best Local Similarity 7.4%; Score 35.8; DB 4; Length 303;
Matches 67; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 84 AGAAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 143
Db 134 AGCCGCTGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 143
OY 144 AGGTAGTGGTGAATGCAATGCAATGCAATGCAATGCAATGCAAT 193
Db 194 AGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 202
RESULT 4


```

US-08-556-978B-81
: Sequence 81, Application US/08556978B
: Patent No. 6268169
: GENERAL INFORMATION:
: APPLICANT: FAHNESTOCK, STEPHEN F.
: TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
: TITLE OF INVENTION: SPIDER SILK ANALOGS
: NUMBER OF SEQUENCES: 107
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
: STREET: 1007 MARKET STREET
: CITY: WILMINTON
: STATE: DELAWARE
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 19898
: COMPUTER READABLE FORM:
: MEDIUM TYPE: DISKETTE, 3.50 INCH
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: MICROSOFT WINDOWS 95
: SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/556, 978B
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/077, 600
: FILING DATE: JUNE 15, 1993
: ATTORNEY/AGENT INFORMATION:
: NAME: FLOYD, LINDA AXAMETHY
: REGISTRATION NUMBER: 33,692
: REFERENCE/DOCKET NUMBER: CR-9389-A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 302-892-8112
: TELEFAX: 302-773-0164
: INFORMATION FOR SEQ ID NO: 81:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 303 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-556-978B-81

Query Match 7.4%; Score 35.8; DB 4; Length 303;
Best Local Similarity 56.3%; Pred. No. 0.032;
Matches 67; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 84 AGAAGTCGAGCTGCGACTGGGAGCTGTGTGGCGCGGTGCTGCTGGCCAGCTTTGGTAA 143
Db 83 AGCCCTCGCGCAGCTGGTGTGCTGAGGCGCGCTCTGGCTCAAGAGGCGCGGTCA 142
OY 144 AGGTAGTCGTCAGTGTGCAATGAGCCATTTGGTGTCTGTTGGGTGATTAATGGTT 202
Db 143 AGGCCCTGTCAGCAGCAGCAGCTGCCCTGCGCGGTGCAGAGCCAAAGTGTGATGTCCT 201

RESULT 5
US-08-556-978B-82
: Sequence 82, Application US/08556978B
: Patent No. 6268169
: GENERAL INFORMATION:
: APPLICANT: FAHNESTOCK, STEPHEN F.
: TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
: TITLE OF INVENTION: SPIDER SILK ANALOGS
: NUMBER OF SEQUENCES: 107
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
: STREET: 1007 MARKET STREET
: CITY: WILMINTON
: STATE: DELAWARE
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 19898
: COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/556, 978B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/077, 600
FILING DATE: JUNE 15, 1993
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9389-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-556-978B-82

Query Match 7.3%; Score 35.4; DB 4; Length 303;
Best Local Similarity 53.2%; Pred. No. 0.043;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0

QY 62 GTGCCAGAACTTATGTCGTCAGAAAGTCGGAGCTGCGACTGGGGCTGTGTTGGCGGTG 121
    ||| |||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 68 GTGGCCAGGTCAGAGTGCTGCTGCTGCTGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTC 127
QY 122 TTGCTGGCCAGCTGTTTGTAAGTAGTGTGTGAGTTGTGCATGGCCATTGGTGTGTGTC 181
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128 TGGGATCTCAGAGGGCCAGAGTCAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 187
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 182 TTTTGGTGTGATTAATGTGTT 202
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 188 GTCAAGGTGGCTACGGCGGTT 208
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
PCT-US95-10668-3/C
Sequence 3, Application PC/TUS9510668
GENERAL INFORMATION:
APPLICANT: James Eberwine
TITLE OF INVENTION: A Method of Sequencing Proteins by
TITLE OF INVENTION: Epitope Ordering and Protein
TITLE OF INVENTION: Restriction Mapping
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jane Massey Licata, Esq.
STREET: 210 Lake Drive East, Suite 201
CITY: Cherry Hill
STATE: NJ
COUNTRY: USA
ZIP: 08002
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10668
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/294,133
FILING DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:

```

us-09-677-374-1 mi

```
Query Match          7.2%; Score 35.2; DB 5; Length 198;
Best Local Similarity 49.5%;
Matches 91; Conservative 0.0; Mismatch 0.0;
```

RESULT 7
PCT-US95.

10

Query Match	7.28;	Score 35.2;	DB 5;	Length 198;
Best Local Similarity	49.58;			
Matches	91;	Conservative	0	

RESULT 8

(103) 836-2021

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1046 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
IMMEDIATE SOURCE:
CLONE: 3C9
US-08-484-332C-4

Query Match          7.1%; Score 34.4; DB 1; Length 1046;
Best Local Similarity 52.9%; Pred. No. 0.18;
Matches    74; Conservative   0; Mismatches    66; Indels     0; Gaps      0.

QY   84 AGAAGTCGACGCTCGACATGGGGGTGTTGTTGGCGGTGCTGTGCACACTGTTTGGTAA 143
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   329 AGGACGACGAATAAGGTGATGGTGTTGGTGGTGGTTTAACCTGGGCTTGACGTGGG 270
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY   144 AGGTAGTGGTCAGATTGCAATGACCATTGGTGCTGCTGTTTGGGNGCATTTAATGGTTC 203
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   269 TTTTGATGGCTATTAAGTGGGAGACCTTAGTGGTGGTTTAACCGGTGCTTTAACTGGTGG 210
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY   204 TAAATCGGTCATCGATGATG 223
      | | | | | | | | | |
Db   209 TTTCGATGGCTATTAAGTGG 190
      | | | | | | | | | |

RESULT 10
PCT-US95-10668-1/c
Sequence 1, Application PC/TUS9510668
GENERAL INFORMATION:
APPLICANT: James Eberwine
TITLE OF INVENTION: A Method of Sequencing Proteins by
TITLE OF INVENTION: Epitope Ordering and Protein
TITLE OF INVENTION: Restriction Mapping
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jane Massey Licata, Esq.
STREET: 210 Lake Drive East, Suite 201
CITY: Cherry Hill
STATE: NJ
COUNTRY: USA
ZIP: 08002
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10668
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/294,133
FILING DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: PENN-0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 198
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: No

```


Tue Oct 29 08:30:52 2002

us-09-677-374-1.rni

ADDRESSEE: CUSHMAN DAREY & CUSHMAN, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D. C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/332,766A
FILING DATE: 01-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9326052.9
FILING DATE: 21-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: BIRD, Donald J.
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 21/211/M94/0434/GB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-332-766A-1

Query Match
Best Local Similarity 6.8%; Score 33; DB 2; Length 377;
Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 78 TCGTCAAGAGTGGAGCTGCGACTGCGGCTGTGTGGCGGCTGTGCGCCAGCTGTT 137
DB 66 TCCCTCGTAAACAAATGGTATGATGATGATGATGATGATGATGATGATGATGATG 137
QY 138 TGGTAAAGTATGCTGCACTTGCATGCGCATGCGTGTGCTGCTGTTGGGGATTAAT 125
DB 126 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 125
QY 196 TGGTTCCTAA 206
DB 186 TGGTGTGCA 194

Search Completed: October 27, 2002, 18:37:15
Job time : 44.5337 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 14:35:03 ; Search time 123.369 Seconds
(without alignments)
6762.541 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486
Sequence: 1 atgaacagagatgtttgca.....aagtcattcaacagaaaa 486

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	486	100.0	486	22	AAF86246	OspA 17kD antigen
2	486	100.0	489	22	AAH79040	Piscirickettsia sa
3	295	60.7	483	22	AAF86247	DNA sequence of E.
4	295	60.7	768	22	AAF86248	DNA sequence of ci
5	73.2	15.1	110	22	AAF86252	PCR primer #2 used
6	71.6	14.7	118	22	AAF86254	PCR primer #4 used
7	69	14.2	102	22	AAF86255	PCR primer #5 used
8	62	12.8	94	22	AAF86253	PCR primer #3 used
9	56.2	11.6	454	20	AAZ06832	Rickettsia 17 kD o

C	10	46.8	9.6	110	22	AAF86256	PCR primer #6 used
C	11	44	9.1	309	23	ABL13399	Drosophila melanog
C	12	44	9.1	2309	23	ABL13298	Drosophila melanog
C	13	42.6	8.8	790	11	AAQ03870	H. influenzae DNA f
C	14	42.6	8.8	850	9	AAAB0227	Sequence of Haemop
C	15	40.4	8.3	1509	23	ABL20279	Drosophila melanog
C	16	40.4	8.3	5162	23	ABL20278	Drosophila melanog
C	17	39.8	8.2	2226	23	AA578586	DNA encoding novel
C	18	39.8	8.2	2292	21	AA578583	Arabidopsis thaliana
C	19	39.6	8.1	522	22	ABA49593	Human breast cell
C	20	39.6	8.1	522	22	ABA67501	Human foetal liver
C	21	39.6	8.1	522	22	ABA34580	Probe #13046 for g
C	22	39.6	8.1	522	22	AAK15928	Human brain expres
C	23	39.6	8.1	522	22	AAK41670	Human bone marrow
C	24	39.6	8.1	522	22	AA122424	Probe #1257 for g
C	25	39.6	8.1	522	22	AA147715	Probe #16401 used
C	26	39.6	8.1	522	22	AA108110	Probe #8101 used t
C	27	39.4	8.1	1395	23	AA589524	DNA encoding novel
C	28	39	8.0	350	22	ABA72229	Human foetal liver
C	29	39	8.0	350	22	ABA38104	Probe #16570 for g
C	30	39	8.0	350	22	AAK20651	Human brain expres
C	31	39	8.0	350	22	AAK46795	Human bone marrow
C	32	39	8.0	350	22	AA125744	Probe #15677 for g
C	33	39	8.0	350	22	AA152634	Probe #21320 used
C	34	39	8.0	600	22	ABA52634	Human foetal liver
C	35	39	8.0	600	22	ABA28224	Probe #6690 for ge
C	36	39	8.0	600	22	AAK07956	Human brain expres
C	37	39	8.0	600	22	AAK33826	Human bone marrow
C	38	39	8.0	600	22	AA116547	Probe #6480 for ge
C	39	39	8.0	600	22	AA139552	Probe #8238 used t
C	40	39	8.0	801	23	AA567681	DNA encoding novel
C	41	39	8.0	8045	21	AA65171	Neurospora crassa
C	42	38.8	8.0	887	22	ABA51240	Human breast cell
C	43	38.8	8.0	887	22	ABA69248	Human foetal liver
C	44	38.8	8.0	887	22	ABA63164	Probe #14630 for g
C	45	38.8	8.0	887	22	AAK17542	Human brain expres

ALIGNMENTS

RESULT 1	AAF86246	standard; DNA: 486 BP.
ID	AAF86246	
XX	AAF86246:	
AC	11-JUL-2001	(first entry)
XX		
DE	OspA 17kD antigen gene.	
XX		
KW	Poikilothermic fish; piscirickettsia salmonis; rickettsial pathogen;	
KW	vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;	
KW	SRS; ds.	
XX		
OS	Piscirickettsia salmonis.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..486
FT		/tag- a
FT		/partial
FT		/product- "OspA"
FT		/note- "Genus specific 17kDa antigen, the sequence does not include a stop codon"
XX	CA2281913-A1.	
PN	17-MAR-2001.	
XX		
PD	17-SEP-1999.	99CA-2281913.
XX		
PR	17-SEP-1999.	99CA-2281913.
XX		


```

OY 301 GATACAGCAATAGTATTAGTGTGAGCCAGTCGACTTACACAGGCTTACATAAGCAA 360
DB 301 GATACAGCAATAGTATTAGTGTGAGCCAGTCGACTTACACAGGCTTACATAAGCAA 360
OY 361 GAGCGTCGCCAGCAATATTGTGCAATTTTCAGCAAAAGCGATGATTCGAGGCGAGAA 420
DB 361 GAGCGTCGCCAGCAATATTGTGCAATTTTCAGCAAAAGCGATGATTCGAGGCGAGAA 420
OY 421 CAAGACATTTAGCGCACTGCATCCCGCAACCGGATGCTGTTGGCAAGTCATTTCACA 480
DB 421 CAAGACATTTAGCGCACTGCATCCCGCAACCGGATGCTGTTGGCAAGTCATTTCACA 480
OY 481 GAAAAA 486
DB 481 GAAAAA 486

RESULT 3
AAF86247
ID AAF86247 standard; DNA: 483 BP.
AC AAF86247;
XX
XX
DT 11-JUL-2001 (first entry)
XX
DE DNA sequence of E. coli optimised ospa gene 17E2.
XX
XX
KW Polkiothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KM vaccine; OSPA; salmonid rickettsial septicaemia; rickettsial disease;
SRS; 17E2; ds.
XX
OS Piscirickettsia salmonis.
OS Synthetic.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..483
FT /tag- a
FT /product- "ospa"
FT /note- "Genus specific 17kDa antigen, the sequence does
not include a stop codon"
XX
XX
PN CA2281913-A1.
XX
XX
PD 17-MAR-2001.
XX
XX
PF 17-SEP-1999; 99CA-2281913.
XX
XX
PR 17-SEP-1999; 99CA-2281913.
XX
PA (KAWW/) KAY W W.
PA (BURJ/) BURIAN J.
PA (KUZV/) KUZV M A.
XX
XX
PI Kay WW, Burian J, Kuzv MA;
XX
XX
DR MPI: 2001-316844/34.
XX
XX
DR P-PSDB; AAB81127.
XX
XX
PT Method for protecting polkiothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the OSPA protein of Piscirickettsia salmonis
XX
XX
PS Example 3; Fig 4C; 35pp; English.
XX
XX
CC This invention relates to a method for the protection against infection
CC of a polkiothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed OSPA, or an immunogenic fragment of
CC OSPA in the form of a vaccine. The method is used for protecting animals,
CC particularly polkiothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid

```

```

CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents DNA which has been optimised for expression in
CC Escherichia coli to encode the P. salmonis OSPA protein. An OSPA protein
CC with an N-terminal fusion partner is used in a vaccine to create an
CC anti-OSPA antibody response.
XX
SQ Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;

Query Match 60.7%; Score 295; DB 22; Length 483;
Best Local Similarity 76.0%; Pred. No. 2,3e-79;
Matches 364; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

OY 8 GAGAGATTGTCGAAAGTAGTACTAATTATTCAGCTGCTTTTACGTGGCGGCC 67
DB 5 GTGTGTCCTGCGAGGCGAGCTCTGTATCAATATCTGTCTGTTCTGCGGTGGCC 64
OY 68 AGAATTTTACTGCTCAAGAGTCGAGCTGCGAGCTGCGGCTGTTGGCGGTGCTG 127
DB 65 AGAATTTTCAAGCCGCGAGAGAGTTGGCGCGCGCAACCGGTGCGGTGCGGTGCG 124
OY 128 GCCAGCTGTTGTTAAAGTAGTGTGTCAGTGCATGCGCATGTTGCTGTTTGG 187
DB 125 GCCAGCTGTTGTTAAAGCTGCTGCTGTGTCATGCGCATGCGGTGCGGTGCTG 184
OY 188 GTGATTAATTGGTGTCAAAATGCGTCAATCGATGGATCGACGAGATTAATTAAGCTAA 247
DB 185 GCGGTCTGATTGCTCTTAATATCGTCAAGCATGGCGACGAGATTAATCAACTGA 244
OY 248 ACCAGATTTTGAAGAGTAAAGCAGGCGCAAGTGCACCTTGCGCTTAATCAGATACG 307
DB 245 ACCAGTCTGGAAGAAAGTAAAGCCGCGCAGGTACTCTGTCGCGTAATCCGACACCG 304
OY 308 GCAATAGTTTATGTTTGAAGCCAGTGTCTTACACAGCTTTCATTAACAGAGCGTC 367
DB 305 GTAAACAGCTACTGTTGGAACCGGTTGCGCACTACAGCGCTTCAACAAACAGGAACGCC 364
OY 368 GCCAGCAATATTGTCGGAATTCAGCAAAAGCGATGATGGAGGCGCAGACAGAGA 427
DB 365 GTACAGCATACTGCGCGGAATTTTCAGCAGAAAGCATGATGCGAGTCAAGAACAGGAAA 424
OY 428 TTTACGGCACTGCATGCGGCAACCGGATGCTGTGGCAAGTCATTTCACAGAAAAA 486
DB 425 TCTACGGCACCGGCGCTCAGCCGATGGCGGTGCGAGTGATGATGACCGAAAAA 483

RESULT 4
AAF86248
ID AAF86248 standard; DNA: 768 BP.
AC AAF86248;
XX
XX
XX
DT 11-JUL-2001 (first entry)
XX
XX
DE DNA sequence of c17e2 ospa construct with N-terminal fusion partner.
XX
XX
KW Polkiothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KM vaccine; OSPA; salmonid rickettsial septicaemia; rickettsial disease;
SRS; 17E2; fusion construct; ds.
XX
XX
OS Piscirickettsia salmonis.
OS Synthetic.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..768
FT /tag- a
FT /product- "c17e2 ospa with N-terminal fusion partner"
FT /note- "No stop codon is given"
XX
XX
FT misc_feature 1..285
FT /tag- b
FT /note- "DNA encoding undefined N-terminal fusion partner"
XX
XX
FT misc_feature 286..768
FT /tag- c

```

/note="Optimised OspA construct c17E2"

CA2281913-A1.

17-MAR-2001.

17-SEP-1999; 99CA-2281913.

17-SEP-1999; 99CA-2281913.

(KAYW/) KAY W W.

(BURI/) BURIAN J.

(KUZV/) KUZV M A.

KAY W W, Burian J, Kuzv M A;

WPI; 2001-316844/34.

P-PSDB; AAB81128.

Method for protecting poikilothermic fish against salmonid rickettsial diseases comprising administering a vaccine containing the OspA protein of *Piscirickettsia salmonis*.

Example 4; Fig 5; 35pp; English.

This invention relates to a method for the protection against infection of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia salmonis*. The method comprises administering an immunogenic amount of *P. salmonis* specific antigen termed OspA, or an immunogenic fragment of OspA in the form of a vaccine. The method is used for protecting animals, particularly poikilothermic fish, against the bacterial pathogen *P. salmonis*. The method is also useful for protecting against salmonid rickettsial septicæmia (SRS) and other rickettsial diseases. The present invention represents *P. salmonis* OspA DNA termed C17E2 optimised for expression in *Escherichia coli* fused to DNA encoding an undefined N-terminal fusion partner. The protein encoded by this fusion construct is used in a vaccine to create an anti-OspA antibody response.

Sequence 768 BP; 220 A; 183 C; 190 G; 175 T; 0 other;

Query Match

Best Local Similarity 76.0%; Score 295; DB 22; Length 768;

Matches 364; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

```

OY 8 GAGCATTTTGGCAAGTAGTACTAATTAATATCATGTCTGTTTAACTGGCTGTCGCC 67
DB 290 GTGCTTCCCGCAGGCGCAGCTCTGATCATTAATCTGTTTCCCTGGTGGTGGCC 349
OY 68 AGAATTTAGTCTCAAGAGTCCGAGCTGCGAGTGGGGCTGTTGGCGGTTGCTG 127
DB 350 AGAATTTAGTCTCAAGAGTGGCGGCGGCGGCGGCGGCTGTTGGCGGTTGCTG 409
OY 128 GCGAGCTGTTGGTAAGGTAGTGTGCTGAGTTCGATGGCATTTGGTGTGTTTGG 187
DB 410 GCGAGCTGTTGGTAAGGTAGTGTGCTGAGTTCGATGGCATTTGGTGTGTTTGG 187
OY 188 GTGATTAATTTGGTCTCAATCGTCAATGATGATGATGATGATTAATAAGGCTTA 247
DB 470 GCGGTCTGATTTGGCTTAATCGTCAAGATGATGATGATGATGATTAATAAGGCTTA 247
OY 248 ACCAGAGTTTGAAGAAGTAAAGCAAGGCAAGTACAGTGGGCTTAATCCAGTACAG 307
DB 530 ACCAGAGTTTGAAGAAGTAAAGCAAGGCAAGTACAGTGGGCTTAATCCAGTACAG 307
OY 308 GCAATTAATTTAGTGTGAGCGAGTCTGACTTACAGCGTTCATTAAGCAAGAGGAGCC 649
DB 590 GCAATTAATTTAGTGTGAGCGAGTCTGACTTACAGCGTTCATTAAGCAAGAGGAGCC 649
OY 590 GTACACCTACTCTGTGAGACCGGTTCCGACCTACACGTTTACAAACAAGAGAGCC 367
DB 590 GTACACCTACTCTGTGAGACCGGTTCCGACCTACACGTTTACAAACAAGAGAGCC 367
OY 368 GCGAGCAATTTTGTGAGATTTTCAAGAAAGGCGATGTTGAGGCGCAAGAGAGAGA 427
DB 650 GCGAGCAATTTTGTGAGATTTTCAAGAAAGGCGATGTTGAGGCGCAAGAGAGAGA 427
OY 428 TTACGCGCATCTGATGCGCGCAACCGGATGCTGTTGCGCAAGTCAATTTCAACGAGAAA 486

```

```

DB 710 TCTAGGCAACCGCGTCCCTCAGCGGATGCGCGTGTGAGAGTATTAACACGAGAAA 768

```

RESULT 5

AAAF86252

AAAF86252 standard; DNA; 110 BP.

AAAF86252;

11-JUL-2001 (first entry)

PCR primer #2 used in cloning an optimisation of OspA gene.

Poikilothermic fish; *Piscirickettsia salmonis*; rickettsial pathogen;

SRS; 17E2; fusion construct; PCR primer; ss.

Piscirickettsia salmonis.

CA2281913-A1.

17-MAR-2001.

17-SEP-1999; 99CA-2281913.

17-SEP-1999; 99CA-2281913.

(KAYW/) KAY W W.

(BURI/) BURIAN J.

(KUZV/) KUZV M A.

KAY W W, Burian J, Kuzv M A;

WPI; 2001-316844/34.

Method for protecting poikilothermic fish against salmonid rickettsial diseases comprising administering a vaccine containing the OspA protein of *Piscirickettsia salmonis*.

Example 3; Fig 4B; 35pp; English.

This invention relates to a method for the protection against infection of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia salmonis*. The method comprises administering an immunogenic amount of *P. salmonis* specific antigen termed OspA, or an immunogenic fragment of OspA in the form of a vaccine. The method is used for protecting animals, particularly poikilothermic fish, against the bacterial pathogen *P. salmonis*. The method is also useful for protecting against salmonid rickettsial septicæmia (SRS) and other rickettsial diseases. The present invention represents *P. salmonis* OspA gene. The OspA gene is used in the method of the invention.

Sequence 110 BP; 13 A; 28 C; 45 G; 24 T; 0 other;

Query Match

Best Local Similarity 15.1%; Score 73.2; DB 22; Length 110;

Matches 87; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

```

OY 56 TTGGCTGTGCGCAGAACTTGTCTCAAGAGTCCGAGCTGCGAGTGGCGGTTGTTG 115
DB 1 TGGCTGTGCGCAGAACTTGTCTCAAGAGTCCGAGTGGCGGCGGTTGTTGTTG 115
OY 116 GCGGTGTGCGCAGAACTTGTCTCAAGAGTCCGAGTGGCGGCGGTTGTTGTTG 165
DB 61 GCGGTGTGCGCAGAACTTGTCTCAAGAGTCCGAGTGGCGGCGGTTGTTGTTG 165

```

RESULT 6
AAAF86254/c
ID AAF86254 standard; DNA; 118 BP.

```

AC AAF86254;
XX
XX 11-JUL-2001 (first entry)
DT
XX PCR primer #4 used in cloning an optimisation of Ospa gene.
DE
XX
XX Polkillothermic fish: Piscirickettsia salmonis; rickettsial pathogen;
KM vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS; 17E2; fusion construct; PCR primer; ss.
XX
OS Piscirickettsia salmonis.
XX
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
XX (BURI/) BURIAN J.
XX (KUZ1/) KUZIK M A.
XX
XX Kay WW, Burian J, Kuzyk MA;
XX
XX WPI; 2001-316844/34.
XX
XX Method for protecting polkillothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of Piscirickettsia salmonis -
XX
XX Example 3; Fig 4B; 35pp; English.
XX
XX This invention relates to a method for the protection against infection
CC of a polkillothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkillothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the
CC invention.
XX
XX Sequence 118 BP; 21 A; 32 C; 28 G; 37 T; 0 other;
SQ
Query Match 14.7%; Score 71.6; DB 22; Length 118;
Best Local Similarity 75.4%; Pred. No. 5.9e-12;
Matches 89; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
OY 211 GGTCAATCGATGATCAGCAGATAAAATAAGCTTAACCAAGCTTTGAAAAAGTTAAAA 270
DB 118 GGTGACAGCATGAGCAGCAGATTAATAATCAAACTGAACAGTCTCGAAAAAGTGAAA 59
OY 271 GCAGGCGCAAGTGCACAGTTGGCGTAATCCAGATACAGGCAATAGTTAGTTGACAC 328
DB 58 GCGGCGCAGGTACTCGTTGGCGTAATCCGACACCGGTAACAGCTACTCTGTGGAAC 1

```

```

KM SRS; 17E2; fusion construct; PCR primer; ss.
XX
XX Piscirickettsia salmonis.
XX
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
XX (BURI/) BURIAN J.
XX (KUZ1/) KUZIK M A.
XX
XX Kay WW, Burian J, Kuzyk MA;
XX
XX WPI; 2001-316844/34.
XX
XX Method for protecting polkillothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of Piscirickettsia salmonis -
XX
XX Example 3; Fig 4B; 35pp; English.
XX
XX This invention relates to a method for the protection against infection
CC of a polkillothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkillothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the
CC invention.
XX
XX Sequence 102 BP; 14 A; 23 C; 30 G; 35 T; 0 other;
SQ
Query Match 14.2%; Score 69; DB 22; Length 102;
Best Local Similarity 80.2%; Pred. No. 3.4e-11;
Matches 81; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
OY 337 ACTTACCAGCGTTTCAATTAAGCAAGACCGTCGCCAGCAATATTTCGAGAAATTTCAGCA 396
DB 102 ACCTACCAGCGTTTCAACAAACAGAAACGCGTCAGCAGTACTGCCGAAATTTCAGCAG 43
OY 397 AAGCGCATGATTGCGAGGCGAAGCAAGAGATTACGGCAC 437
DB 42 AAMGCCATGATCGCAGGTCCAGAAACAGGAATTTACGGCAC 2

```

```

RESULT 7
AAF86255/C
ID AAF86255 standard; DNA; 102 BP.
XX
XX AAF86255;
XX
XX 11-JUL-2001 (first entry)
DT
XX PCR primer #5 used in cloning an optimisation of Ospa gene.
DE
XX Polkillothermic fish: Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;

```

```

RESULT 8
AAF86253
ID AAF86253 standard; DNA; 94 BP.
XX
XX AAF86253;
XX
XX 11-JUL-2001 (first entry)
DT
XX PCR primer #3 used in cloning an optimisation of Ospa gene.
DE
XX Polkillothermic fish: Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS; 17E2; fusion construct; PCR primer; ss.
XX
XX Piscirickettsia salmonis.
XX
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX

```

Example 3; Fig 4B; 35pp; English.

Query Match	Score	DB	Length
Best Local Similarity	62	22	94
Matches 74; Conservative	78.78		
	0	4.3e-09	
		Mismatch	

```
RESULT 9
AAZ06832
ID      AAZ06832 standard; DNA; 454 BP
AC      xxx
xx      AAZ06832;
```

Key	Location/Qualifiers
FH	
FT	
ET	
CDS	1 454

```

L1      /tag= a
L2      /partial
FT      /product= "17 kd outer membrane protein"
FT      /note= "Stop codon not given in specification"
FT      277..360
FM      /tag= b
FM      /product= "28 amino acid antigenic peptide from
XX      17 kd OMP"
XX      MO9942479-A1.
XX

```

New peptide from outer membrane protein of *Rickettsia helvetica* used to diagnose rickettsiosis and in vaccines

Query Match Post 1168 138 A; 84 C; 117 G; 115 T; 0 other.

	82	CAAGAACTCGACCTGCGACTGGGCCTGTTCATTGGCCGCTTTCCTCCTTTATCCTTGCTTGA	Indels	3;	Gaps	1,
by						
	ac					

85 AAGGTACAGAAACACTCTTGGGCGGCGCGCGGCGATTCATTCTTGGTTCTCAATTTGGT 144
 QY 142 AAGAGTAGTGTCGAGTTGCATATGGCCATTGGTGTGTCTGTTTGGGTGATTAATTGGT 144
 Db 145 AAGGTAAAGGGCAACTTGTTCGGAGTAGGTGTAGGTGCATTACTTGGACAGCTTTGGCG 204
 QY 202 TCTAAATCGTCATCGATGATGGATCAGCAGGATTAATAAGCTAAACCAAGTTGGAA 261
 Db 205 GGGCAAACTGTTGGCAGGTATGATGAGCAGGATGAAGACTTGCAGACCTTACCTCACAG 264
 QY 262 AAGGTAAAGCAGGAGCAATGACACAGTTGGCGTAAATCCAGATACAGCAATATGTTAAT 321
 Db 265 AGACCTTTAAGAC ---AGCTCTTACGCGGTAGTAAAGTAGAGTGGCCTATTCGGGATAC 321
 QY 322 GTTCAGCCAGTCGCTACTTACACAGCTTAACTAAAGCAAGAGCGTCGCCAGCAATATTTGT 381
 Db 382 CGGCAATTTTCAGCAAAAGGAGATGATTCGAGGCGCAGAGCAAGATATTTTACGCAATATTCG 381
 QY 382 CCGAGTACACTCAACACAGTTGTAATAGCGGAAACACAAAAGCATCGGTAATGCA 441
 QY 442 TGCCTGACGACGG 434

XX		31-AUG-1989;	89MO-US03779.
PF			
XX		01-SEP-1988;	88US-0239572.
PR		21-AUG-1989;	89US-0396572.
XX		(PRAK-) PRAXIS BIOLOGICS IN.	
PA		Anilionis A, Seid RC, Deich RA,	Zlotnick GW, Green BA;
XX		WPI, 1990-115815/15.	
DR		P-PADB; AAR05799.	
XX			
PS		Outer membrane protein epitopes of Haemophilus influenzae used in	
PT		diagnosis.	
XX			
PS		Disclosure; Fig 14; 164pp; English.	
CC		The sequence is a 789bp BstEII-XmlI fragment of PAJ30.The predicted	
CC		coding region is given above.	
CC		See also AAQ03869,R03948 and AAR03949.	
XX			
SQ		Sequence 790 BP; 240 A; 106 C; 176 G; 266 T; 2 other;	
		Query Match	
		Best Local Similarity 8.8%; Score 42.6; DB 11; Length 790;	
		Matches 84; Conservative 0; Mismatches 69; Indels 0; Caps	
Oy		75 TAGTCGACAGAGTGGCGAGCTGCACACTGGGCGTTGTGGCGGTGGTGCCAGCT	
Dd		DB 334 TAATCAGGTTAGTGTGTACGCTTGTGTGGAGCTTAGTGTAATGCTGTAGTAC	
Oy		135 GTTGTGAAGGTAGTGTGCAATGCCAATGGCATGGTGTCTCTTTTGGGTGATT	
Dd		394 AATGTGCGGTGCTGTGTGTGCAAGCTTATTCACAGCACTAGTGTGTGCAATGGCGTGCAT	
Oy		195 AATGTGCTTAAATCCGTCATATCGATGATGATCA 227	
Dd		454 AGCTGAGATAAATCGAACAAAAAATGATGCA 486	
		RESULT 14	
ID		AAN80227 standard; DNA; 850 BP.	
XX		AAN80227;	
AC			
XX		12-JAN-1991 (first entry)	
DT			
DE		Sequence of Haemophilus influenzae Praxix Biologics Outer	
DE		Membrane Protein-2 (PBOMP-2) gene.	
KW		Vaccine; diagnosis; epitope; passive immunisation; ss.	
XX		Haemophilus influenzae.	
OS			
XX			
FH		Key	
FT		CDS	Location/Qualifiers
FT			164..628
XX			/*tag= a
PN		WO8804932-A.	
XX			
PD		14-JUL-1988.	
XX			
PF		23-DEC-1987;	87MO-US03423.
XX			
PR		11-DEC-1987;	87US-0132073.
PR		02-MAR-1987;	87US-0020849.
PR		31-DEC-1986;	86US-0948364.
FA		(PRAK-) PRAXIS BIOLOGICS IN.	
XX			

```

PI Delch RA, Zlotnick G, Green B:
XX
DR WPI: 1988-205305/29.
DR P-PSDB: AAP80667.
XX
PT Pure peptide related to epitope of Haemophilus influenzae -
PT used as immunogens in vaccines and for producing antibodies for
PT passive immunisation and assays
XX
PS Disclosure: Fig 14; 129pp: English.
XX
CC A pure antigenic peptide or protein related to an epitope
CC of Haemophilus influenzae is claimed. Also claimed is a
CC recombinant vector comprising a DNA sequence coding for an
CC antigenic determinant of an Haemophilus influenzae outer membrane
CC protein, the transformed cell, a subunit vaccine in a pharmaceutical
CC carrier, a method of immunising humans and an assay for Haemophilus
CC influenzae.
XX
SQ Sequence 850 BP; 254 A; 116 C; 190 G; 288 T; 2 other:
XX
Query Match 8.8%; Score 42.6; DB 9; Length 850;
Best Local Similarity 54.9%; Pred. No. 0.0088;
Matches 84; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 75 TAGTCGTCAGAGTCCGAGCTCGAGCTGGGGCTGTGTGGCCGTGTGTCGCCAGCT 134
DB 394 TAATCAAGGTGTAGTGTGTACGCTTGCTGGAGCTTTAGTGTGTATTTGCTGTATTAC 453
QY 135 GTTTGGTAAAGGTAGTGTGTCGATGCAATGGCCATTGTGTGCTGTTTGGGTGAT 194
DB 454 AATTGGGGGTGTCGTGTGTCAAGCTATTGCACGAGTAGTGTGCAATTGGCGGTCAAT 513
QY 195 AATTGTTCTAAATTCGTCATCGATGATCA 227
DB 514 AGCTGGAAGTAAATCGAAGAAAATAGATCA 546

RESULT 15
ABL20279/c
ID ABL20279 standard; DNA: 1509 BP.
XX
AC ABL20279;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 12310.
XX
KM Drosophila: developmental biology; cell signalling; insecticide;
KM pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN W0200171042-A2.
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 12310; 21pp + Sequence Listing; English.

```

```

XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB37737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 1509 BP; 495 A; 446 C; 277 G; 291 T; 0 other:
XX
Query Match 8.3%; Score 40.4; DB 23; Length 1509;
Best Local Similarity 49.1%; Pred. No. 0.053;
Matches 107; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
QY 8 GAGCATGTTTCCAGAGTAGTACTAATTATATCATCAGTCTTTTACTTGGCTGTGCC 67
DB 985 GTGGTCTGTGGTGTGCTGTCTGTGGAGGTTCTGTGTGGTGTGCGTCCGTCG 926
QY 68 AGACTTTAGTCCTCAAGAACTCGAGCTCGAGCTGGGGCTGTGTGGCGGTGTGCTG 127
DB 925 GTGGCTTTGTGTGTGCTGTACTTGGAGGAATTGTGGAGTAATTGTGGAGGAATTGTTG 866
QY 128 GCCAGCTGTTTGTGTAAGGTAGTGTGCAATGGCCATTGTGTGCTGCTTTTGG 187
DB 865 GAGCAATTGTGAAGGAATTGTGTGAGAGATTTGTATGGCTTGTGTGTGACTAGTCG 806
QY 188 GTGATTAATTGTTCTAAATTCGTCATCGATGAT 225
DB 805 GGGGAAGTGTGGGTGATTAAGTAGTGTGATGTGTTGGT 768

Search completed: October 27, 2002, 16:40:39
Job time : 127.389 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 15:07:23 : Search time 941.223 Seconds
(without alignments)
10805.408 Million cell updates/sec

Title: US-09-677-374-1
Perfect score: 486
Sequence: 1 atgaacagagagatgttgca.....aagtcattcaacagaaaa 486

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pal:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description

1	486	100.0	489	6	AX252413	AX252413 Sequence
2	486	100.0	4983	1	AF184152	AF184152 Piscirickettsia
3	79	16.3	448	1	RSU76907	RSU76907 Rickettsia
4	70.6	14.5	237523	1	RPRX04	RPRX04 Rickettsia
5	68.4	14.1	416	1	AF031534	AF031534 Rickettsia
6	65.8	13.5	535	1	RIRANT17KB	RIRANT17KB Rickettsia
7	65.8	13.5	537	1	RIR17KCA	RIR17KCA Rickettsia
8	64.6	13.3	539	1	RIRANT17KA	RIRANT17KA Rickettsia
9	64.6	13.3	10127	1	AE008675	AE008675 Rickettsia
10	64.2	13.2	539	1	RIRANT17KD	RIRANT17KD Rickettsia
11	64.2	13.2	620	1	RIRANT17	RIRANT17 Rickettsia
12	62.2	12.8	434	1	MRI269516	MRI269516 Male-Kill
13	62.2	12.8	434	1	MRI269517	MRI269517 Male-Kill
14	62.2	12.8	434	1	MRI269518	MRI269518 Male-Kill
15	59.4	12.2	546	1	AF195118	AF195118 Rickettsia
16	59	12.1	489	1	RPU17008	RPU17008 Rickettsia
17	58.2	12.0	415	1	AF027124	AF027124 Rickettsia
18	57.8	11.9	532	1	RIRANT17KC	RIRANT17KC Rickettsia
19	57.6	11.9	411	1	AF060704	AF060704 Rickettsia
20	57.6	11.9	411	1	AF060706	AF060706 Rickettsia
21	57.4	11.8	489	1	RPU17008	RPU17008 Rickettsia
22	57.4	11.8	489	1	RAU11013	RAU11013 Rickettsia
23	56.2	11.6	454	1	AF181036	AF181036 Rickettsia
24	55.6	11.4	489	1	U04162	U04162 Rickettsia
25	54.4	11.2	395	1	U04162	U04162 Rickettsia
26	54.2	11.2	492	1	RIRRAPRO	RIRRAPRO Rickettsia
27	52.6	10.8	491	1	RIRGENSEO	RIRGENSEO Rickettsia
28	47.8	9.8	394	1	AF260571	AF260571 Rickettsia
29	47.8	9.8	7218	6	166494	166494 Sequence 14
30	47.4	9.8	63454	2	AC102004	AC102004 Mus muscu
31	46.4	9.5	111824	9	AL160412	AL160412 Human DNA
32	45.4	9.3	68650	9	AC027343	AC027343 Homo sapi
33	45.4	9.3	123670	2	AC103214	AC103214 Rattus no
34	45	9.3	156370	2	AL158162	AL158162 Homo sapi
35	45	9.3	166464	9	AL158212	AL158212 Human DNA
36	44.6	9.2	13123	1	AE003872	AE003872 Xylella f
37	44.6	9.2	46877	9	AC104133	AC104133 Homo sapi
38	44.4	9.1	184999	2	AL611934	AL611934 Mus muscu
39	44.2	9.1	474	3	PFU2952	PFU2952 Plasmodiu
40	44.2	9.1	475	3	AF104714	AF104714 Plasmodiu
41	44.2	9.1	585	3	PFAMS21	PFAMS21 Plasmodiu
42	44.2	9.1	615	3	PFAMS21	PFAMS21 Plasmodiu
43	44	9.1	68418	2	AC018090	AC018090 Drosophila
44	44	9.1	153048	3	AC008362	AC008362 Drosophila
45	44	9.1	175302	2	AC020587	AC020587 Homo sapi

ALIGNMENTS

RESULT 1
AX252413
LOCUS AX252413 489 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 5 from Patent WO0168865.
AX252413
ACCESSION AX252413
VERSION AX252413.1 GI:15985721
KEYWORDS
SOURCE
ORGANISM
Piscirickettsia salmonis.
Piscirickettsia salmonis.
Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;
Piscirickettsia.
1 (bases 1 to 489)

REFERENCE
AUTHORS Simard,N., Brouwers,H., Jones,S.F., Griffiths,S., Valenzuela,P. and

TITLE Fish vaccine against piscirickettsia salmonis
JOURNAL Patent: WO 0168865-A 5 20-SEP-2001;
Aqua Health (Europe) Limited (GB)
Location/Qualifiers

FEATURES
source

BASE COUNT 139 a 79 c 144 g 127 t
ORIGIN

CDS

0;

```
<1, .501
/gene="alr"
/codon_start=1
```

gene
complement(526. .1257)
/gene="Bax"
complement(526. .1257)
/gene="Bax"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="BAX"
/protein_id="AAC17002.1"
/db_xref="GI:1041348"
/translation="MELSLAVYITLVAADQPLFSAPTSTKTOITN
KLTSDIIPNIIITLKKRHITLSLQKSSISKEITLOALAKPTKLN
MEWANSMTSTSVATITLNTSNMWSRSPVAEGNNFGMCHPLGG
GKRYVELTQITITLITLTKSESTQIP"
/gene="rad4"
/evidence=not_experimental
1373. .2743

gene
2834. .3322

CDS	/evidence=experimental
2834	3333

```

/codon_start=1
/transl_table=11

```

```

/db_xref="GI:10441346"
/translation="Protein"

```

2. ONSGR VAMALIGGAVLGLIGSKIGÓSMDOQDKIKL NÓSL EKVKAGOVTEBUDVAGG
 GNSYSEVPARTYORYNKOEPPOOVCTTUEV
 3.

```

/strain="LF-89"
/dh vnf="

```

repeat_region	insertion_seq="Insertion sequence-like element
gene	3560. .3816
	complement(3842. .4561)
CDS	/gene="trpA"
	complement(3842. .4561)

Query Match	Best Local Similarity	100.0%;	Score 486;	DB 1;	Length 4983;
Matches 486;	Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	ATGAAACAGAGATGTTTTCAGACGTAAGTGTATATATATATACAGTGTGTTTGTAGTGGC	60			
Db 2834	ATGAAACAGAGATGTTTTCAGACGTAAGTGTATATATATATACAGTGTGTTTGTAGTGGC	2893			
QY 61	TGTGCCCCAGAACTTTAGTCGTCAGAAAGTGCAGACTGCAGCTGGGGCGTTGGTGGCGGT	120			
Db 2894	TGTGCCCCAGAACTTTAGTCGTCAGAAAGTGCAGACTGCAGCTGGGGCGTTGGTGGCGGT	2953			
QY 121	GTTCGTGGCCAGCTGTTTGTGTAAGTAGTGTGAGTTCGAATGCGCATTGGTGTGCT	180			
Db 2954	GTTCGTGGCCAGCTGTTTGTGTAAGTAGTGTGAGTTCGAATGCGCATTGGTGTGCT	3013			
QY 181	GTTTGTGGGTGATTAATTTGCTTTCTAAATCGGCTCAATCGATGATCGACAGATTAATA	240			
Db 3014	GTTTGTGGGTGATTAATTTGCTTTCTAAATCGGCTCAATCGATGATCGACAGATTAATA	3073			
QY 241	AAGCTAAACAGAGATTTGGAAGAGTAAGCAAGGCAAGTGCAGACGTTGGCGTAATCCA	300			
Db 3074	AAGCTAAACAGAGATTTGGAAGAGTAAGCAAGGCAAGTGCAGACGTTGGCGTAATCCA	3133			
QY 301	GATACAGGCAATAGTTATAGTGTGAGCCAGTGCCTACTTACACAGCGTTACATTAAGCAA	360			
Db 3134	GATACAGGCAATAGTTATAGTGTGAGCCAGTGCCTACTTACACAGCGTTACATTAAGCAA	3193			
QY 361	GAGCGTCCGCCAGCAATATTTGTCGAGAAATTTTACGAAAAAGCGATGATTTGCAGGGCAGAG	420			
Db 3194	GAGCGTCCGCCAGCAATATTTGTCGAGAAATTTTACGAAAAAGCGATGATTTGCAGGGCAGAG	3253			
QY 421	CAAGAGATTTACGGCACTGCATGCCGCGCAACCGGATGGTCTTGGCAAGCATTTTCAACA	480			
Db 3254	CAAGAGATTTACGGCACTGCATGCCGCGCAACCGGATGGTCTTGGCAAGCATTTTCAACA	3313			
QY 481	GAAGAAA 486				
Db 3314	GAAGAAA 3319				
RESULT 3	RSU76907	448 bp	DNA	linear	BC1 18-FEB-1998
LOCUS	RSU76907				
DEFINITION	Rickettsia sp. 17kda common-antigen gene, partial cds.				
ACCESSION	U76907				
VERSION	U76907.1				
KEYWORDS	GI:2894794				
ORGANISM	Rickettsia sp.				
SOURCE	Rickettsia sp.				
REFERENCE	1 (bases 1 to 448)				
AUTHORS	Davis,M.J., Ying,Z., Brunner,B.R., Pantoja,A. and Ferwerda,F.H.				
TITLE	Rickettsial relative associated with papaya buncy top disease				
JOURNAL	Curr. Microbiol. 36 (2), 80-84 (1998)				
MEDLINE	98087556				
REFERENCE	2 (bases 1 to 448)				
AUTHORS	Ying,Z. and Davis,M.J.				
TITLE	Direct Submission				

[illegible]

MEDLINE

99039499
2 (bases 1 to 237523)
REFERENCE
AUTHORS
TITLE
JOURNAL

Andersson, S.G.E.
Submitted
Siv Andersson (11-NOV-1998)

FEATURES

of Upsala, Husaratan 3, Uppsala, S-751 24, SWEDEN
Location/Qualifiers
1. 237523

source

/organism="Bickettsia prowazekii"

/strain="Madrid E"

/db_xref="taxon:782"

/gene="Rp701"

/gene="Rp701"

/codon_start=1

/transl_table=1

/product="unknown"

/protein_id="CA15137.1"

/db_xref="GI:3861238"

/translation="MAPIITFTLLFGNKRDLIRNNKAKANKLHIIPLQVKNESHIS
ALPQNSLSEKSKYKQIIEKVDHVEFTNEADONLSIADGIVPKDRTTIDILVSVN
EIOYIGIKLEHILKDPNNAAENFKKYSOISROKPKDLDKDNHIDSNFNQAKIT
NDSNKDIOVEHITSYVDKREKOSKPGFLIRIPNFKNIVKRO"

/gene="Rp702"

2074. 3348

/gene="Rp702"

/codon_start=1

/transl_table=1

/product="unknown"

/protein_id="CA15138.1"

/db_xref="GI:3861239"

/translation="MTTLLVVIIMISLAFATETATASSBQKIHKLKAGNR
VPAEVPKAIARPEOLILKNSLINTICTTATTFTLSLNGGTIVASGVMAFTI
ISGTEVKGVIIEHYHOEGVYKSEKRNMGITIDIRNMTSEIMTHRSNLIANDP
ELITITLLSGAHTRIPLMONRDNITGLADIRNMTSEIMTHRSNLIANDP
ETEDNAVFQDLAHRNNHACVVDDEYGLIGITLEDVIEETVGPIDDEHRLNN
ELIQSNTEFTIKCTTIRIDINRELMDSTLADITGLITLEDVIEETVGPIDDEHRLNN
NKKITILKTIANKIDSKITVDPMEETISSE"

/gene="Rp703"

3341. 5353

/gene="Rp703"

/codon_start=1

/transl_table=1

/product="CITTOCHROME C-TYPE BIOGENESIS PROTEIN CCMF
(ccmp)"

/protein_id="CA15139.1"

/db_xref="GI:3861240"

/translation="MSKIGNFLITMCISGLITLLIIPMDYLKKHRRNICLKE
SOSMSHEGMLIMFCLLOVSCYFECLEDMALFESSILISAQILFSSFTYFN
PHEMSVEPGLGIMQDIALSTHPLDGLGVYAVPPTVCAIFLAVISRD
NLPDGYFTGYKITQISRTISRLIKIFAGCGIETTGISLGSMAVRELNMGLMF
NSIHAFASSESEWITVLVLCIGIGLVYIRLNKNSIERKTIINEEFTKNTIIPVILIC
TIGIFTNSSIKRHIMLSLITFTPLINIKRSVLSITMTSIFLMHVVILY
KTNVPRNSIKRHIMLSLITFTPLINIKRSVLSITMTSIFLMHVVILY
LONIKRQKRNYYQIAEFMLEDKNNITTLKPENNRYIIEOKLSESDIHSYLIYDL
YALSLIDKRIIHAARYKPKMSFTIGVLTITAGVIALIKRNSY"

/gene="Rp704"

complement(7264. 12195)

/gene="Rp704"

/codon_start=1

/transl_table=1

/product="CELL SURFACE ANTIGEN (scas5)"

gene

complement(7264. 12195)
/gene="Rp704"

CDS

/codon_start=1
/transl_table=1
/product="CELL SURFACE ANTIGEN (scas5)"

misc-feature

/note="Pseudogene that lacks ATG or GTG initiation codon.
3-phosphoglycerolase."

/gene="Rp705"

complement(12953. 13216)

/gene="Rp705"

complement(13563. 14507)

/gene="Rp705"

complement(13563. 14507)

/gene="Rp706"

/codon_start=1

/transl_table=1

/product="unknown"

/protein_id="CA15141.1"

/db_xref="GI:3861242"

/translation="MKIRQAPVITIGSELTDAERELFEHNPLGITLERNIRK
TCKYNSITAKELREGLINDEPAVADILHGAOKIISDSRKEEIVYPLFELA
MFAHVITLDPDNPATLSQIVIDYIKSNIGPGILITSDIEQALSGSMADITKCAL
DACVIDYLTCTEENKMEVIGSAENSIDKPSDLITS"

/gene="Rp707"

complement(14552. 15025)

/gene="Rp707"

/codon_start=1

/transl_table=1

/product="unknown"

/protein_id="CA15142.1"

/db_xref="GI:3861243"

/translation="MINNKLKKYVAVSEVIRIINTALHKLRLKIRGLSHYKINN
KNSKELINLHVQRODOSAVVLSLMDNTYSRIDILLINHLNLSIOIKKILLS"

/gene="Rp708"

complement(15220. 15507)

/gene="Rp708"

/codon_start=1

/transl_table=1

/product="INTEGRATION HOST FACTOR ALPHA-SUBUNIT (hima)"

/protein_id="CA15143.1"

```

/db_xref="GI:3861244"
/translation="MTITKNKISIMLSKIGFENN/CFAIVNTVPSNILEIKVOKLT
LNKSGFEVKOKTPRGINFPHRTKSPMIAKNNLRTPSEKILKALINKSMU"
gene
18403..18701
/gene="RP709"
18405..18701
/gene="RP709"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA15144.1"
/db_xref="GI:3861245"
/translation="MAVROYIEDHTTEOLHLCDNLVKYLEETKAITHEKAEDTEV
VIEVSLDINSDDASIIKKKAITVSDSIANSNEDANNTKSHODVLRCTYQ"
gene
complement(19685)..20789
/gene="RP710"
complement(19685)..20789
/gene="RP710"
/feature="Pseudogene with two authentic frameshifts and four
authentic in-frame termination codons. Homology to
transposase."
gene
complement(21112)..21608
/gene="RP711"
complement(21112)..21608
/gene="RP711"
/feature="Pseudogene with one authentic frameshift. Homology
to resolvase."
misc_feature
complement(22602)..23066
/gene="RP712"
complement(22602)..23066
/gene="RP712"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA15145.1"
/db_xref="GI:3861246"
/translation="MSIERKEFMASHKQVKEIGTGNLNGKHDLFKAVAYTSDIEA

Query Match      14.5%: Score 70.6; DB 1; Length 237523;
Best Local Similarity 50.6%; Pred. No. 2.3e-08;
Matches 197; Conservative 0; Mismatches 189; Indels 3; Gaps 1;

QY 82 CAAAGATGCGAGCTGGAGCTGGGGCTGTGTGGCGGTGTGGCCAGCATGTTGGT 141
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172624 CAAAGATGCGAGCTGGAGCTGGGGCTGTGTGGCGGTGTGGCCAGCATGTTGGT 172683
QY 142 AAAGTAGTGGTGAAGTGGCCATGTTGTTGCTGTTTGGGTGATTAATTTGT 201
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172684 CAAAGTAGTGGTGAAGTGGCCATGTTGTTGCTGTTTGGGTGATTAATTTGT 172743
QY 202 TCTAAATGCGTCAATCGATGATCAGCAGGATTAAGCTAAACCGAGTTTGGAA 261
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172744 GGAACAATCGCGCAGTGAAGTGAAGCAGATGAGACATTTCTAGAACTAACATCAAA 172803
QY 262 AAGGTAAACGAGGCGAAGTGAAGCAGTGGCGTAAATCCAGATACAGATTAATAGT 321
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172804 AAGCTTTAGAAATCTGCACTACGCGTA---GTAACTAAATAGGCGCAATCCAGATTAAC 172860
QY 322 GTTGAGCGACTGCGTACTTACACAGCGTTTACAAATAGCAGAGCGTCCGCAATATTTGT 381
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172861 GGCATATCATGTTAGCTACACACCTTAATAAACTTATAGAAACAGTGCAGGTCAATATTGC 172920
QY 382 CGAGATTTTCAGAAAAAGGCGATGATTTGCGAGCGCAGAGAGATTTTACGCGACATGCA 441
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172921 CGTGAATACACTCAACAGTTATTAATAGGCGGAAACACAAAAAATATGTGTAATGCA 172980
QY 442 TGCCGCGAACCGGATGCTGTGGCAAGT 470
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172981 TGCCGCGAACCGGATGCTGTGGCAAGT 173009

```

```

LOCUS      AF031534      416 bp      DNA      linear      BCT 09-JAN-1998
DEFINITION Rickettsia coolleyi 17 kDa antigen gene, partial cds.
ACCESSION  AF031534
VERSION    AF031534.1 GI:2654012
KEYWORDS   Rickettsia coolleyi.
SOURCE     Rickettsia coolleyi.
ORGANISM   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsiinae; Rickettsia.
REFERENCE  1 (bases 1 to 416)
AUTHORS    Billings,A.N., Tellow,G.J. and Walker,D.H.
TITLE      Molecular characterization of a novel spotted fever group
            rickettsial species from Ixodes scapularis in Texas
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 416)
AUTHORS    Billings,A.N., Tellow,G.J. and Walker,D.H.
TITLE      Direct Submission
JOURNAL    Submitted (28-OCT-1997) Pathology, UTMH at Galveston, 301
            University Blvd., Galveston, TX 77555-0609, USA
FEATURES   location/Qualifiers
            source
            1..416
            /organism="Rickettsia coolleyi"
            /db_xref="taxon:69410"
            <1..415
            /codon_start=2
            /transl_table=11
            /product="17 kDa antigen"
            /protein_id="AB95267.1"
            /db_xref="GI:2654013"
            /translation="GSYIGPRGMNKGDTLLGAGGALLGSGFGKGOIVGVGVA
            LGAVLGGIAGMDEODRRLAELTSORLEAPSGSSTEMRRPDNGNYVPNNY
            RNSGYCKREYQTVIVIGGKQKQAKVGNACRQPEQ"
BASE COUNT 133 a 80 c 108 g 95 t
ORIGIN
Query Match      14.1%: Score 68.4; DB 1; Length 416;
Best Local Similarity 49.5%; Pred. No. 4.1e-08;
Matches 205; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

QY 51 TTTAGTGGCTGTGCCAGAACTTTAGTCGTCAAGAAAGTCGAGCTGCGGCTGT 110
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4 TTCAATATATCGGTCACGCGGTATGATTAACAAAGCAGCAACCTCTTGGCGGTGC 63
QY 111 TGTGGCGGCTGTGCTGCCAGCTGTTGGTAAAGTATGATGCTGCAATGGCCAT 170
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 64 CCGCGGTGCAATTCCTTGCGTCTCAATTCGTTAAAGGTAAGGACAGCTTTCGAGTAGG 123
QY 171 TGGTGTGCTGTTTGGGTGATTAATTTGTTCAATATGCTCAATCATGATGATCAGCA 230
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 124 TGTAGTGCATTAATCTTGACAGAGTCTTGGTGGCAAAATCGGTGCAGATGATGACGA 183
QY 231 GGATAAATTAAGCTAAACAGAGTTTGGAAAGTAAAGCAAGGCAAGTGCACGCTTG 290
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184 GGAATAGAGACTTGCAGAACCTCACTCACAAGAGCTTTAGAG---CACCTCTTAGCGG 240
QY 291 GCGTAATCCAGATACAGAGCAATAGTTATAGTTGAGACCGAGTGTACTTAACAGCGTTA 350
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 TAGTAGTACAGAAATGCGGTATCTGTATTAAGGCAATTAACGTTATTAACACCTAATAA 300
QY 351 CAATTAACAGAGCGTGGCCAGCAATATTTGCGAATTTTCAGAAAAAGCGATATTCG 410
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 AACTTATAGAAATATAGCACTGTGCAATATTTGCCGAGATATATCTCAACACAGTTGTAATAGG 360
QY 411 AGGCGAGACAGACAGATTTACGCGACTGCATGCGCGGCAACCGGATGCTGTTG 464
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 CGGAAACCAACAAAGACATACGCTAATGATGATGCGCGCCACACCTGACGAAACAATG 414

```

RESULT 5
AF031534

RESULT 6
RIRANT17KB
LOCUS RIRANT17KB 535 bp DNA linear BCT 26-APR-1993
DEFINITION R. typhi 17k genus-common antigen gene, complete cds.
ACCESSION M28481

```

-35-signal
-10-signal
mRNA
CDS
/db_xref="taxon:785"
1..6
21..27
37..>535
/product="17 kd mRNA"
56..535
/note="17 kd antigen precursor"
/codon_start=1
/transl_start=1
/protein_id="AA26377.1"
/db_xref="GI:152460"
/translation="MLLSKYVILALASLTQNCPEGNNKQGTGLLAGAGALD
SQFSGHGGLGVYVGGVALLGVLLGQIGASLDGPRKLELTTSORALESPGSGNTEW
RPNPNHGHCYVFNPNNTYNSGYCBEYVQYVIGCKQKQITTYGNACRQDPCQWQVNN"
56..115
/ncore="17 kd antigen signal peptide"
116..532
/product="17 kd antigen"
93 c 122 g 138 t

```

	Query Match	13.5%	Score 65.8;	DB 1;	Length 535;
	Best Local Similarity	49.9%;	Pred. No. 2.2e-07;		
	Matches 194;	Conservative	0;	Mismatches 192;	Gaps
QY	82 CAAAGATCGGAGCTGGAGCTGGGGCTGTTGTGGCGGTCTTCCTGCCAGCTGTTGGT	1111	1111	1111	1111
Db	140 CAAAGGACTGGACACACTCTTGGTGGCCCGCGGTGCATTACTGGTGTCCAAATTCGG	1111	1111	1111	1111
QY	142 AAAGTACTGGTGCAGTTCGAATGGCCAAATGGGTGGCTGTTGGGTGAATATGGT	1111	1111	1111	1111
Db	200 CACGGTAAAGGACAACTTCTCGAGTAGTGTTAGGGCACTTACTGGGGCAGTTCTTGGT	1111	1111	1111	1111
QY	202 TCTAAATCGGTCAATCGATGATGACACAGATAAATAAAGCTAAACAGAGTTGGCA	11111111	111111111111	11111111	11111111
Db	260 GGACCAATCGGTGCAAGTCTGATGAGCCAGATAGAAACCTTCTAGCAATCAATCAACA	11111111	111111111111	11111111	11111111
QY	262 AAGGTAAAGCAGGGCAAGTAGACACCTTGGCTATTCAGTAGACAGCAATGTTATGCT	11111111	111111111111	11111111	11111111
Db	320 AGAGCTTAGAATCT---GCTCCTACGGTAGTAGCACTAAAGAGCGCAATCGAATATAT	11111111	111111111111	11111111	11111111
QY	322 GTTGGCCAGTGGCTACTTACCAAGCGTTACAAATAGACAGAGGGTGGCCACCAATATGT	11111111	111111111111	11111111	11111111
Db	377 GGCATATATGTTACGTAAACCTTAAATAACTTATAGGAACAGTACAGGTCAATATTTG	11111111	111111111111	11111111	11111111
QY	382 CGAATATTCAGCAAAAGCCATGATTTGCAGGGCCAAACAGCAAGATTTAGGGCACTGCA	11111111	111111111111	11111111	11111111
Db	437 CGTGAATACACTTAAACAGTTGTAAATGGGGGAAACCAACAACAATATATGGAATATGA	11111111	111111111111	11111111	11111111
QY	442 TGGCGCAACCGGATGCTGTTGGCAAGT	11111111	111111111111	11111111	11111111
Db	497 TGGCGCAACCTGAAGGACCAATGGCAAGT	11111111	111111111111	11111111	11111111

RESULT 7
RIR17KGCA

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
LOCUS	R1R1/KGCA	537 bp	DNA	linear	BCN 04-FEB-1995	
DEFINITION	<i>Rickettsia japonica</i> 17k genus-common antigen gene, complete cds.					
ACCESSION	D16515					
VERSION	D16515.1	GI:695412				
KEYWORDS	17k genus-common antigen, 17k dalton protein.					
SOURCE	<i>Rickettsia japonica</i> (sub__species:yH) DNA.					
ORGANISM	<i>Bacteria</i> ; <i>Proteobacteria</i> ; <i>alpha</i> subdivision; <i>Rickettsiales</i> ; <i>Rickettsiaceae</i> ; <i>Rickettsia</i> ; <i>Rickettsia</i> sp. 1 (bases 1 to 537)					
REFERENCE	Yoshida, Y.					
AUTHORS	Submitted (05-JUL-1993) Yoshida Yoshida, Kanagawa Prefectural Public Health Laboratory, 55-2 Nakao-cho, Asahi-ku, Yokohama, Kanagawa 241, Japan (Tel:045-363-1030, Fax:045-363-1037)					
TITLE	Furuya, Y., Katayama, T., Yoshida, Y. and Kalho, I.					
JOURNAL	Specific amplification of <i>Rickettsia japonica</i> DNA from clinical specimens by PCR					
FEATURES	J. Clin. Microbiol. 33 (2), 467-469 (1995)					
source	Location/Qualifiers					
	I..537					

```

/organism="Rickettsia japonica"
/sub_species="YH"
/db_xref="taxon:35790"
/note="propagated in Vero cells"
58      537
/feature="unnamed protein product"
/codon_start=1
/transl_table=11
/translation="KLLSKIMIIATLSWLAQNPQGNKKQSTGILGAGAGALLIC
SOFGKNGTGVGALGAVLGGQIGAGAGBDDRIALETTSQRALETAPGSSWVEH
RNFENGTGVTPNPTVNSTGVCYCRFYDTDTVIGGKKQAKYAGACRODPCQKQV
129      140

```

	Query Match	Best Local Similarity	Score	DB 1;	Length	537;
	Matches	194, Conservative	49.98;	Pred No. 2,2e-07;	Mismatches	192;
					Indels	3;
					Gaps	
Oy	82	CAGAACTCGAGCTCGACTGGGCTGTTGGCGGTGGTGGCCAGCTTTGGT	13.5%;	DB 1;		
Db	142	CAAGGTACAGAAACCTCTTGGCGGTCTGGTGGCCAGCTTTGGT	49.98;			
Oy	142	AAAGTAGTGGTCGAGTTGCATAGGCCATTGGTGTCGCTTTTGGGTGGATTAATGGT				
Db	202	AAAGGACAGAGACAGCTTGGCGAGTAGTGTAAGTGAATGACTTACTTGGAGCAGTCTTGGT				
Oy	202	TCTAAATCGTCAATGATGATCGATCGACGAGATAAATTAACCTTAACCAAGCTTTGGAA				
Db	262	GGACAAATCGGTGCAGCTATGATGATGGACAGGATGAAGAAGACTTGCAGACCTTTACTCTACAG				
Oy	262	AAGTAAACACAGGGCAAGTGAACAGCTTGGGCTTAATCGAGTACAGGCAATATTCTTAAT				
Db	322	AGAGCTTTAAGAAC--ACGCTCTAGTGTGTGTAACGTGAAGTGGGTATATCGGATAC				
Oy	322	GTTGACCGCAGTCCGTACTTACCAGCGTTACATTAAGCAAGCAGCGTCCGACAGTAATTTGT				
Db	379	GGCAATTACGGTTAGCTTAACACCTAATTAATAACTTAATAGAATATGACATGTGTAATAATTTGC				
Oy	382	CGAAGATTTTCAGCAAAAAGCGAGTATTCAGAGGCGACAGACAGATTTTACGCGACTGCA				
Db	439	CGTGTACTACTCAAAACAGTTGTAAATAGCGGGAACCAACAAAAAGATACGGTAATCGG				
Oy	442	TGCCGGAACCGGATGGTGGTTGGCAAGT				
Db	499	TGCGGCAACCTGACGACATATGGCAAGT				

RESULT 8

RIRANT17KA 539 bp DNA linear BCT 26-APR-1993
 LOCUS RIRANT17KA
 DEFINITION R.conorii 17K genus-common antigen gene, complete cds.
 ACCSSION M28480
 VERSION M28480.1 GI:152457
 KEYWORDS antigen.
 SOURCE R.conorii DNA.
 ORGANISM Rickettsia conorii
 Bacteria: Proteobacteria; alpha subdivision; Rickettsiales;
 Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
 REFERENCE 1 (bases 1 to 539)
 AUTHORS Anderson,B.E. and Tzanabos,T.
 TITLE Comparative sequence analysis of a genus-common rickettsial antigen gene
 JOURNAL J. Bacteriol. 171, 5199-5201 (1989)
 MEDLINE 89359171
 COMMENT Draft entry and printed copy of sequence for [1] kindly submitted by B.Anderson, 27-OCT-1989.
 FEATURES
 source
 1..539
 /organism="Rickettsia conorii"
 /db_xref="taxon:781"
 1..6
 -35_signal
 21..27
 38..>539
 /product="17 kd mRNA"
 60..539
 /note="17 kd antigen precursor"
 /codon_start=1
 /transl_table=11
 /protein_id="AA26376.1"
 /db_xref="GI:152458"
 /translation="MKLSKIMIIATSMLOACNPGCMKNGTGLLGAGAGALLG
 SQPKRGQLVGVGALGAVIGGIGAGMDQDRLELTSQRLLETPASSNVEW
 RPNDSNGYIVPNKTYRNSYGVYCREYQTVVIGKQKAVGNACRODPGQWVNV"
 60..119
 /note="17 kd antigen signal peptide"
 120..536
 /product="17 kd antigen"
 93 c 125 g 142 t
 BASE COUNT 179 a 93 c 125 g 142 t
 ORIGIN
 sig_peptide
 mat_peptide
 Query Match 13.3%; Score 64.6; DB 1; Length 539;
 Best Local Similarity 49.4%; Pred. No. 4.7e-07;
 Matches 197; Conservative 0; Mismatches 199; Indels 3; Gaps 1;
 72 CTTTAGTCGTCAAGAACTGCGAGCTGCGGCTGTTGCGGCTTGGCCCA 131
 134 CATGAATTAACAAGGTACAGGACACTTCTGGCGGTGCGGCCCATTTACTTGTTC 193
 132 GCCTTTGGTAAAGGTAGTGTGCAATGGCCATTGGTGCTGTTTGGGTGG 191
 194 TCATTCGGTAAGGGCAAGGACGTTGTGGAGTAGGTGCTGCTTACTTGGAGC 253
 192 ATTAATTGTTCTTAATCGTCAATGATGATCAGCAGATTAATAAAGCTAAACCA 251
 254 AGTCTTGGTGGCAAAATCGGTGAGATGATGAGAACAGATGAAAGACTGTGAGAGCT 313
 252 GAGTTTGGAAAAGGTAAAGCAGGGCAAGTGACAGCTTGGCGTAATTCAGATACAGGCCAA 311
 314 TACCTCACAGAGAGCTTTAGAAAC--AGCTCTAGTGTAGTAAAGTAGAATGGCGTAA 370
 312 TAGTTATAGTGTGAGCACTGCGTACTTACACAGCTTACAAATTAAGCAAGAGCTGCCCA 371
 371 TCCGGATTAAGGCAATTACGTTACGTAACACCTAATTAATAACTTATGAAATATGCACTGG 430
 372 GCAATATTGTGAGAAATTTGAGCAAAAGCGCATGATGAGCGGAGAGCAAGAGATTTA 431
 431 TCAATATTGCGCGTGTAGTACACTCAAAACAGTTGTATATAGCGGAAAACACAAAAAGCATA 490
 432 CGGCATGTATGCCGCGGACACCGATGCTGTGGCAAGT 470
 491 CGGTAATGATCATGCCGCGCACCTGACGACATATGCAAGT 529

RESULT 9
 LOCUS AE008675
 DEFINITION AE008675 10127 bp DNA linear BCT 14-SEP-2001
 Rickettsia conorii Malish 7, section 107 of 114 of the complete genome.
 ACCSSION AE008675 AE006914
 VERSION AE008675.1 GI:15620413
 KEYWORDS Rickettsia conorii.
 SOURCE Rickettsia conorii.
 ORGANISM Bacteria: Proteobacteria; alpha subdivision; Rickettsiales;
 Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
 REFERENCE 1 (sites)
 AUTHORS Ogata,H., Audic,S., Barbe,V., Artiguenave,F., Fournier,P.E.,
 Raoult,D. and Claverie,J.M.
 TITLE Selfish DNA in protein-coding genes of Rickettsia
 JOURNAL Science. 290 (5490), 347-350 (2000)
 MEDLINE 20485642
 PUBMED 11030655
 REFERENCE 2 (sites)
 AUTHORS Ogata,H., Audic,S. and Claverie,J.-M.
 TITLE Selfish DNA and the origin of genes
 JOURNAL Science 291 (5502), 252-253 (2001)
 REFERENCE 3 (bases 1 to 10127)
 AUTHORS Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V.,
 Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. and
 Raoult,D.
 TITLE Mechanisms of Evolution in Rickettsia conorii and R. prowazekii
 JOURNAL Science. 293 (5537), 2093-2098 (2001)
 PUBMED 11557893
 REFERENCE 4 (bases 1 to 10127)
 AUTHORS Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V.,
 Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. and
 Raoult,D.
 TITLE Direct Submission
 JOURNAL Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine,
 CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean
 Moulin, Marseille Cedex 05 13385, France
 COMMENT A public version of R. conorii genome database is accessible at
 http://igs-server.cnrs-mrs.fr/. The database intends to provide
 updated data. Annotation of the genome is an ongoing task whose
 goal is to make the genome sequence more useful. Comments to the
 authors are appreciated.
 FEATURES
 source
 1..10127
 /organism="Rickettsia conorii"
 /strain="Malish 7"
 /db_xref="taxon:781"
 111..890
 /gene="map"
 /note="RC1276"
 111..890
 /gene="map"
 /codon_start=1
 /transl_table=11
 /product="methionine aminopeptidase [EC:3.4.11.18]"
 /protein_id="FAL03814.1"
 /db_xref="GI:15620414"
 /translation="MTIKIHGTGDFIKMRAAGLAAETLDFITDHVKNPTVNSLGL
 CHNFIRSHNAVPAPIVYKGPSPKSTICSHNVYCHGIPNPKPLKNGDIVINDVVIIDG
 WNGDTRMYIVGDVAIKPKRLIDYTTIDAMAKGIEVYRPPAKLADISHAOSYAEKINY
 SVVRDYGHGIGVHFDEPSILNMGSGTGLTLEEGMFTVEPFINAGNVDTLTKLKD
 GMTVTRDKSLSQPFHTTIGVTGDFEITLSPKLDYPPY"
 complement(977..1011)
 /note="REP03, repeated element"
 complement(1064..1083)
 /note="REP02, RS3-like repeat"
 1141..1452
 /gene="RC1277"
 1141..1452
 /gene="RC1277"
 repeat_region
 repeat_region
 gene
 CDS
 /gene="RC1277"
 /gene="RC1277"


```

Db 8328 AGTTCCTGTGACAAATCGCTCAGGTATGATGAACAGGATAGAACTTCGACAGCT 8387
QY 252 GAGTTTGGAAGAAGTAACAGCAGGGCAAGTACACGCTGGCGTAATTCACATACAGGCA 311
Db 8388 TACCTACACAGAGAGCTTTACAAAC--AGCTCCTAGTGTAGTAACGTAAATGCGCTAA 8444
QY 312 TAGTTATAGTTGAGCCAGTGGCTACTTACACGCTTACAAATGAAGCAAGGCGTCGCA 371
Db 8445 TCCGGATTAACAGGCAATTAACGTTACGTACACCACTTAATAAAGCTTATGAATATGACACTGG 8504
QY 372 GCATATATTGTGAGAAATTTACAGCAAAAGCGAGTATTCAGCCGACAGCAAGAGATTGA 431
Db 8505 TCATATATTCGCGTGAGTACACTCAAAACAGTTGTAATAGCGCAAAACAAACAAAAGCAT 8564
QY 432 CGGACATCGCATCGCGCAACCGGATGCTGTTGGCAAGT 470
Db 8565 CGGTAATGATGCCGCCACCTGACGCAATAGGCAAGT 8603

RESULT 10
RIRANT17KD 539 bp DNA linear BCT 26-APR-1993
LOCUS RIRANT17KD 539 bp DNA linear BCT 26-APR-1993
DEFINITION R.rickettsia 17k genus-common antigen gene, complete cds.
ACCESSION M28479
VERSION M28479.1 GI:152463
KEYWORDS antigen.
SOURCE R.rickettsia DNA.
ORGANISM Rickettsia rickettsii
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE 1 (bases 1 to 539)
AUTHORS Anderson,B.E. and Tzianabos,T.
TITLE Comparative sequence analysis of a genus-common rickettsial antigen
gene
JOURNAL J. Bacteriol. 171, 5199-5201 (1989)
MEDLINE 89359171
COMMENT Draft entry and printed copy of sequence for [1] kindly submitted
by B.Anderson, 27-OCT-1989.
FEATURES
    source
        location/Qualifiers
            1..539
                /organism="Rickettsia rickettsii"
                /db_xref="taxon:783"
    -35_signal
        21..27
    -10_signal
        37..>539
    mRNA
        /product="17 kD mRNA"
        60..539
    CDS
        /note="17 kD antigen precursor"
        /codon_start=1
        /transl_table=11
        /protein_id="AA26379.1"
        /db_xref="GI:152464"
        /translation="MKLSKIMIALATSMLOACNPGMKNKGTTLLGAGGALLG
        SFGKRGOLVGVGALGAVLGGQIGAGMDEODRLAELTSORALETPARSSNEM
        RNPNDNGYVTPNKTYRSTGYCREYQTVIVIGKQKQAVGNACROPDGMQVNV"
    sig_peptide
        60..119
        /note="17 kD antigen signal peptide"
    mat_peptide
        120..536
        /product="17 kD antigen"
BASE COUNT 179 a 92 c 125 g 143 t
ORIGIN
Query Match 13.28; Score 64.2; DB 1; Length 539;
Best Local Similarity 49.68; Pred. No. 6.1e-07;
Matches 193; Conservative 0; Mismatches 193; Indels 3; Gaps 1;

```

```

QY 202 TCTAAATCGTGCAATGATGATCAGCAGATTAATAAGCTAAACAGAGTTGGAA 261
Db 264 GACAAATTCGGTGCAGGTATGATGAATGAACAGATGAAGACTTGCAGCTTACCTCACAG 323
QY 262 AAGTAAACAGCAGGCGCAAGTACACGCTTGGCGTAATTCAGATACAGCAATAGTTATAGT 321
Db 324 AGAGCTTTAACAAC--AGCTCCTAGTGTAGTACGTACAAATGCGGTATCCGGATTAAC 380
QY 322 GTTGAGCCAGTGGCTACTTACACAGCTTACAAATPAGCAAGCGTGGCCAGCAATATTGT 381
Db 381 GGCATATTACGCTTACGTACCACTTAATAAACTTATGAAATATGCACTGGTCAATATTGC 440
QY 382 CGAGATTTACAGCAAAAGCGATGATTCGAGGCGAGAGCAAGATTTACGGCACTGCA 441
Db 441 CGTGAGTACACTCAAAACAGTTGTAATAGCGGAAACACAAACAAAGCATAGGTAATGCA 500
QY 442 TGCCGCGCACCGGATGCTGTTGGCAAGT 470
Db 501 TGCCGCGCACCTGACGCAATAGGCAAGT 529

RESULT 11
RIRANTRR 620 bp DNA linear BCT 26-APR-1993
LOCUS RIRANTRR 620 bp DNA linear BCT 26-APR-1993
DEFINITION R.rickettsii 17k antigen gene, complete cds.
ACCESSION M16486
VERSION M16486.1 GI:152467
KEYWORDS antigen.
SOURCE R.rickettsii (strain Sheila Smith) DNA, clone pSC2.
ORGANISM Rickettsia rickettsii
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE 1 (bases 1 to 620)
AUTHORS Anderson,B.E., Regnery,R.L., Carlone,G.M., Tzianabos,T.,
McDade,J.E., Fu,Z.Y., and Bellini,W.J.
TITLE Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia
rickettsii
JOURNAL J. Bacteriol. 169, 2385-2390 (1987)
MEDLINE 87222152
COMMENT Draft entry and computer-readable sequence for [1] kindly provided
by B.E.Anderson, 23-OCT-1987.
FEATURES
    source
        location/Qualifiers
            1..620
                /organism="Rickettsia rickettsii"
                /db_xref="taxon:783"
    sig_peptide
        70..126
        /note="17kD antigen signal peptide"
        70..549
    CDS
        /note="17kD antigen precursor"
        /codon_start=1
        /transl_table=11
        /protein_id="AA26381.1"
        /db_xref="GI:152468"
        /translation="MKLSKIMIALATSMLOACNPGMKNKGTTLLGAGGALLG
        SFGKRGOLVGVGALGAVLGGQIGAGMDEODRLAELTSORALETPARSSNEM
        RNPNDNGYVTPNKTYRSTGYCREYQTVIVIGKQKQAVGNACROPDGMQVNV"
    mat_peptide
        127..546
        /product="17kD antigen"
BASE COUNT 206 a 105 c 142 g 167 t
ORIGIN
Query Match 13.28; Score 64.2; DB 1; Length 620;
Best Local Similarity 49.68; Pred. No. 6.2e-07;
Matches 193; Conservative 0; Mismatches 193; Indels 3; Gaps 1;

```

```
Query Match      12.88;  Score 62.2;  DB 1;  Length 434;
```

```

/assay="CAB96382.1"
/product="17 kDa antigen"
/protein_id="CAB96382.1"

```

```

/product="I / kDA antigen"
/protein_id="CAB96382.1"
/db_xref="GI:8920295"
/db_xref="SPRREMBL:O9K2N6"

```

primer_bind	415..434	/translation="AATATSMACCTGPEGNKKCGTTLGGAGCAGALLGSOPKRGKGGVGCNVALGAVIGGOTGAGCGEDDDRRPAELTSGRALFAAPSSGNVEMRNPDNNGHGVTPKKTYRNSGTGQCREYTGTVVIGKKQKATGNACRQPDQ"
BASE COUNT	132 a 88 c 114 g 100 t	
ORIGIN		
Query Match	12.8%; Score 62.2; DB 1; Length 434;	
Best Local Similarity	50.1%; Pred. No. 2.1e-06;	
Matches 182; Conservative	0; Mismatches 178; Indels 3; Gaps 1;	
OY	102 TGGGGCTGTTGGCGCGGTGCTGCGCCAGCTGTTGGTAAAGTAGTGGTCGAGTTGC	161
DB	75 TGGCGTCCCGCGGTGATTACTTGTTCTCAATTGCGCAAAAGGTAAAGGCCAATTGT	134
OY	162 AATGSCCATTTGGTGTGCTGTTGGGTGGATTAATTGGTTCTAAATCGGCAATCGAT	221
DB	135 CGGAGTAGTGTAGTGGCATTTACTTGGAGCAGCTTTGGTGGACAAATCGGTGCGAGT	194
OY	222 GGATCAGCAGGATAAATAAAGCTTAACCGAGTTGGAAAAAGTAAAGCAGCGCAAT	281
DB	195 GGATGACGAGATAGAACAGCTTGCGAATCAGCTCGCAAGAGCTTTAGAAAGC--AGC	251
OY	282 CACACGTTGGCCTAATCCAGATACAGCAATGTTATAGTGTGGCCAGTGGCTACTTA	341
DB	232 TCCCTAGGTGTAGTACGTAAATGGCGTAAATCCCATTAACGGCAATCATGTGCTAGCTAAC	311
OY	342 CCAGCGTTAACAATAAGCAAGCGTGGCCAGCAATATGTGAGAAATTTCAAGCAAAAGC	401
DB	312 AACTAATAAACTTATAGAAATAGCACAGGGCAATATTGGCCGTGACTCAAAACAGT	371
OY	402 GATGATTCCAGGCGAGAAAGACAGAGATTTCAGCGCACTGATGCCGCAACGGATGCTCG	461
DB	372 TGTATAGCGGAAAAACAACAAAGCATPAGCGTAAATGCATGCGCCCAACCTGAGAGACA	431
OY	462 TTG 464	
DB	432 ATG 434	
RESULT 14		
LOCUS	MR1269518	
DEFINITION	Male-killing Rickettsia from Adalia bipunctata partial 17kDa antigen gene (Denmark: Ribe).	
ACCESSION	AJ269518	
VERSION	AJ269518.1 GI:8920296	
KEYWORDS	17 kDa antigen; male-killing Rickettsia from Adalia bipunctata; male-killing Rickettsia from Adalia bipunctata; Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia.	
SOURCE	1 (bases 1 to 434)	
REFERENCE	von Der Schulenburg, J.H., Habig, M., Sloggett, J.J., Webberley, K.M., Bertrand, D., Hurst, G.D. and Majerus, M.E. Incidence of male-killing rickettsia spp. (alpha-proteobacteria) in the ten-spot ladybird beetle adalia decempunctata L	
AUTHORS	Appl. Environ. Microbiol. 67 (1), 270-277 (2001)	
TITLE	2 (bases 1 to 434)	
JOURNAL	Schulenburg, H.	
MEDLINE	Direct Submission	
REFERENCE	Submitted (03-APR-2000) Schulenburg H., Department of Evolutionary Biology, Institut fuer Spezielle Zoologie, Huefferstr. 1, 48149 Muenster, GERMANY	
JOURNAL	Location/Qualifiers	
FEATURES	1..434	
source	/organism="male-killing Rickettsia from Adalia bipunctata"/specific_host="Adalia bipunctata"/db_xref="taxon:38028"/country="Denmark: Ribe"/note="Isolated from 2spot ladybird"	

primer_bind	1..20	/note="PCR primer R1"
CDS	<1..>434	
	/transl_start=1	
	/transl_table=1	
	/product="17kDa antigen"	
	/protein_id="CA86383.1"	
	/db_xref="GI:8920297"	
	/db_xref="SPRMBL:O9K2N6"	
	/translation="ALATSMLOACTGCGGNKOGTGLLGAGCAGALLGSGFGKKGQLVPEVGVGLGVLGGOIGAGMEDQDRRLAELTSSRALFEAPSGSNVEMRPNDKNNHYVTPKTKRNSNGYCRREYTVIVIGGKQKRAYGNACNRPDEQ"	
primer_bind	415..434	
	/note="PCR primer R2"	
BASE COUNT	133 a 87 c 114 g 100 t	
ORIGIN		
Query Match	12.8%; Score 62.2; DB 1; Length 434;	
Best Local Similarity	50.1%; Pred. No. 2.1e-06;	
Matches 182: Conservative	0; Mismatches 178; Indels 3; Gaps 1;	
Oy	102 TGGGCGCTGTTGGCGGCGGTGCTGGCGCAGCTGTGTGTTAAGGTAGTGGTCGATTGC	161
Db	75 TGGCGGTCGCGCGGCGTGCATTACTTGTTCTCAATTCGCAAGGTAAAGGCAACTTGT	134
Oy	162 AATGGCCATTGGTGCTGCTGTTTGGGTGGATTAATGTTCTAAATCGGTCAATCGAT	221
Db	135 CGGAGTAGCTGTAGTGGTGCATTACTGGAGCAGTCTTGTGTGACAAATCGGTGCAAGTAT	194
Oy	222 GGATCAGCAGCATAAATAAAGCTTAACCCAGAGTTGGCAAAAGGTAAGGAGGCAAGT	281
Db	195 GGATGAGCAGCATAAACACTTGTGAATCTCAATCTCAAGAGAGCTTTAAGAAC--AGC	251
Oy	282 GACAGCTGGCGGTATCCAGATACAGGAATATGATATGATGTGACGACAGTCCACTTA	341
Db	252 TCTTATGTTGTAACGTAGTAAGTGGCGTAAATCCGCAATTCATGCTACCTTAAC	311
Oy	342 CCAGCGTTACATAAGACAAGAGCGTCGCCAGCAATATTTGTCGAGAAATTCAGCAAAAGC	401
Db	312 ACCTAATAAACTTATATGAAATATAGCACAGGGCAATATTTGCCGTAGTCACTCAAAAGT	371
Oy	402 GATGATTGCGAGGCGCAAGACAGATTTACGGCACTGTCATGCGCGCAACCGGATGTCG	461
Db	372 TGTATATAGCGGAAACAAACAAGACATGCGTAAATCAATCGCGCAACCTGACGAACA	431
Oy	462 TTG 464	
Db	432 ATG 434	
RESULT 15		
AF195118	546 bp	DNA linear BCT 12-JUN-2001
LOCUS	AF195118	
DEFINITION	Rickettsia felis 17 kDa genus-common antigen gene, complete cds.	
ACCESSION	AF195118	
VERSION	AF195118.1	GI:11066087
KEYWORDS		
SOURCE		
ORGANISM	Rickettsia felis.	
	Rickettsia felis.	
	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;	
	Rickettsiaceae; Rickettsiaseae; Rickettsia; spotted fever group.	
REFERENCE	1 (bases 1 to 546)	
AUTHORS	Boyer, D.H., Stenos, J., Crocquet-Valdes, P., Moron, C.G., Popov, V.L., Zavalis-Velazquez, J.E., Foll, L.D., Stothard, D.R., Azad, A.F. and Walker, D.H.	
TITLE	Rickettsia felis: molecular characterization of a new member of the spotted fever group	
JOURNAL	Int. J. Syst. Evol. Microbiol. 51 (Pt 2), 339-347 (2001)	
MEDLINE	21217364	
PUBMED	11321078	
REFERENCE	2 (bases 1 to 546)	
AUTHORS	Boyer, D.H., Stenos, J., Crocquet-Valdes, P.A., Foll, L.D. and Walker, D.H.	

us-09-677-374-1.rge

Direct Submission
Submitted 6/14/2007

TITLE Direct Submission
JOURNAL Submitted (14-Oct-1999) Pathology, University of Texas Medical
Branch at Galveston, 301 University Blvd., Galveston, TX
77551-0609, USA
FEATURES

localtion/qualifiers
1. .546

```

/db_xref="taxon:42862"
65. 544

```

```

/coo_start=1
/transl_table=11
/prod=1

```

```

/protein_id="AAG28452.1"
/db_xref="GI:11066088"

```

SQFGKGGQLVGVGALLGAV
 RNPDNGNHGYVTPNKTVPNGS

```

BASE COUNT      183 a      96 c      126 g      141 t
ORIGIN
RNPONGNHGVTPTPNKTYRNSGQYLSAGMEDODRLRLAELTSSORALEAPSGTSVEW

```

Query Match
Post r

Every match 12.2%; Score 59.4; DB 1; Length 546;
 Best Local Similarity 48.8%; Pred. No. 1.3e-05;
 Matches 190; Conservative 0

[illegible]

D6
149 CAAGGTACGGGAACACTTCTTGAGCGGCGCCGTTGGT 14

142 AAAGTACTGCTCGATTGCAATGGGATTTCTCATTTGGC 208

D6
209 AAGGCAAGACACCTTCCCGCCTTCCTGGTGCTGTCTTTGCCTGATTAATTGGT 201

202 TCTAATAATCGGCTCAACGCTTCTTGGT 268

Ddb
269 GGACAAATAGCGGCTGGTGTGTTGGA 261

263 AACCTAAAGGCTTGGCTGAACTCACTTCACAA 328

320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

323 AGACCTTAGAGCAACACCCTA---CGGCACTAGCGTAGAATGGCGTAATCCGCAATAAG 326

322 GTTGAGCCAGTGCCTACTTACCAGCTTACATATAGCAAGAGGTTGCCACGAAATGTTTAC
323 | | | | |
324 | | | | |

386 GGTAATCATGTTACGTAAACACCTAATAAACCTTATAGGAATAGCAGCGGGTTCCTCCGCCAATATGCT 381

CGAGATTTCAGCAAAAGCGATGTTGCAGGGCAGAACCACCTTCC

CGTGACTACCTCAACAGTTGTAATAGCGGGAAACAACACCCCTTACGGCCTGCA

442 TGGCGCAACGGATGCTGTTGGCAAGT 470
 443 TGGCGCAACGGATGCTGTTGGCAAGT 505

506 TGGCGCCAACTGACGACTATGGCACT 534

Search completed: October 27, 2002, 17:39:09
Job time : 1078.22 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 20:15:49 : Search time 973.679 seconds
(without alignments)
6736.842 Million cell updates/sec

Title: us-09-677-374-1

Perfect score: 486

Sequence: 1 atgaacagggaggtgttgc.....aagtcattccacagaaaa 486

Scoring table: OLIGO_NUC

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:
1: em_estdb:*
2: em_esthm:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_iny:*
15: em_gss_pln:*
16: em_gss_vtl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	4.3	661	9	BB517742
2	20	4.1	395	10	BE428593
3	20	4.1	673	10	BF643947
4	20	4.1	689	12	AZ731780
5	20	4.1	988	12	CNS06GLH
6	19	3.9	376	9	AJ003546
7	19	3.9	380	10	T59966
8	19	3.9	436	9	AM151916
9	19	3.9	461	12	AZ952777
10	19	3.9	550	12	AZ887587
11	19	3.9	545	9	A1965711
12	19	3.9	561	12	BM091630
13	19	3.9	582	10	BM082634
14	19	3.9	584	10	BM083068
15	19	3.9	585	10	BI474188
16	19	3.9	617	10	BM160573
17	19	3.9	723	12	AG132034

C 18	19	3.9	852	12	AQ742455
C 19	19	3.9	1179	10	BG114834
C 20	19	3.9	1295	10	BM006536
C 21	18	3.7	163	9	AI081506
C 22	18	3.7	200	9	AI205976
C 23	18	3.7	200	9	AI205977
C 24	18	3.7	200	9	AI205978
C 25	18	3.7	214	9	AI208086
C 26	18	3.7	276	9	BB416046
C 27	18	3.7	284	9	BB480551
C 28	18	3.7	286	9	BB186207
C 29	18	3.7	359	10	BE949717
C 30	18	3.7	373	10	BE949717
C 31	18	3.7	379	10	BE952725
C 32	18	3.7	407	12	AQ339510
C 33	18	3.7	408	12	AZ119827
C 34	18	3.7	417	10	BE672478
C 35	18	3.7	424	10	BF223697
C 36	18	3.7	436	10	NB1359
C 37	18	3.7	463	10	BE335854
C 38	18	3.7	466	9	AM123846
C 39	18	3.7	468	9	AM221110
C 40	18	3.7	469	10	C87441
C 41	18	3.7	479	10	BG202620
C 42	18	3.7	484	10	BG067771
C 43	18	3.7	485	9	AV847706
C 44	18	3.7	491	9	AM449278
C 45	18	3.7	492	9	AI463386

ALIGNMENTS

RESULT 1
BB517742
LOCUS
DEFINITION
BB517742 RIKEN full-length enriched, 16 days neonate heart Mus
musculus cDNA clone DB30025004 3', mRNA sequence.

ACCESSION
BB517742

VERSION
BB517742.2

KEYWORDS
GI:16443203

SOURCE
Mus musculus

ORGANISM
house mouse.

REFERENCE
Arakawa, T., Carrinchi, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koyu, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, J., Shibata, K., Shimagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M., and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)

On Jul 28, 2000 this sequence version replaced gi:9569200.

Contact: Yoshinori Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carrinchi, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10): 1617-1630 (2000)
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and

Hayashizaki, Y.
RIKEN Integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multichipillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y., and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nondundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamana, I., Aizawa
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., and
Hayashizaki, Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Location/Qualifiers
1. 661
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D830025004"
/clone_lib="RIKEN full-length enriched, 16 days neonate
heart"
/tissue_type="heart"
/dev_stage="16 days neonate"
/lab_host="DH10p"
/note="Site 1: Salt; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer 15'
GAGAGAGAGAGCGCGCCGACACTGAGTTTGTGTTTATTTTNN 3', cDNA was
transcribed using trehalose thermo-activated reverse
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence 15'
GAGAGAGAGATCTCGAGTTAATTAATTAATCCCGCCCGCC 3'. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from lambda FLX 1."

BASE COUNT
ORIGIN
175 a 162 c 171 g 153 t

Query Match
Best Local Similarity 4.3%; Score 21; DB 9; Length 661;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 171 TGGTGCTGCTGTTTGGGTGG 191
|||||
Db 505 TGGTGCTGCTGTTTGGGTGG 525

RESULT 2
BE428593/c
LOCUS BE428593 395 bp mRNA linear EST 26-JUL-2000
DEFINITION MTD008.G05F990617 ITEC MTD Durum Wheat Root Library Triticum
ACCESION BE428593
VERSION BE428593.1 GI:9426436
KEYWORDS EST.
SOURCE durum wheat.
ORGANISM Triticum durum subsp. durum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
1 (bases 1 to 395)
AUTHORS Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier

JOURNAL
COMMENT
TITLE
S. Dubcovsky, J. Feuillet, C. Gale, M. Graner, A. Gustafson, P.
Langridge, P., Lazo, G.R., Jacquemont, J.M., Jia, J., Joudrier, P.,
Pecchioli, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sorrells, M., Warburton, M., and Wenzel, G.
Expressed Sequence Tags for Species of the Triticaceae
Unpublished (2000)
Contact: Joudrier P
INRA, unite de Biochimie et Biologie Moleculaire des Cereales
2, place VIALA, 34060 Montpellier cedex 01 FRANCE
Tel: 33 4 99 61 23 84
Fax: 33 4 99 61 23 48
Email: joudrier@enscm.inra.fr
International Triticaceae EST Cooperative (ITEC)
<http://wheat.pw.usda.gov/genome>.
Location/Qualifiers
1. 395
/organism="Triticum durum subsp. durum"
/cultivar="Silliana"
/db_xref="taxon:4567"
/clone="MTD008.G05"
/clone_lib="ITEC MTD Durum Wheat Root Library"
/tissue_type="root"
/dev_stage="3-day-old seedling, water-stressed"
/note="Vector: pSPORT1; T7 primers used. See pSPORT1
polylinker site. 0.3-2.0 kbp average insert size."

BASE COUNT
ORIGIN
98 a 111 c 114 g 72 t

Query Match
Best Local Similarity 4.1%; Score 20; DB 10; Length 395;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 116 GCGGTGCTGCGCCAGCTG 135
|||||
Db 48 GCGGTGCTGCGCCAGCTG 29

RESULT 3
BE43947/c
LOCUS BE43947 673 bp mRNA linear EST 20-DEC-2000
DEFINITION NF026G01EC1F1005 Elicited cell culture Medicago truncatula cDNA
ACCESION BE43947
VERSION BE43947.1 GI:11909076
KEYWORDS EST.
SOURCE Medicago truncatula
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago
1 (bases 1 to 673)
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Iman, J.T., Weller, J.W., and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Center for Medicago Genomics Research
Unpublished (2000)
Contact: Dixon RA
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert Length: 673 Std Error: 0.00
Plate: 026 row: g column: 01
Seq primer: TCACACGAGAAAGCTATGAC.
Location/Qualifiers
1. 673
/organism="Medicago truncatula"
/db_xref="taxon:3880"

FEATURES
source

/clone="NF026G01EC"
/tissue_type="Elicited cell culture"
/dev_stage="Cell suspensions were subcultured every 14 days. Cells were induced six days after subculture"
/note="Vector: Lambda Zap; Cells were induced with yeast cell wall extracts equivalent to 50ug/ml glucose in the final concentration. Samples were taken at 0.5, 1, 12 and 24 hours after induction. Equal amounts of RNA from each time point were pooled and used for mRNA isolation."

BASE COUNT 175 a 151 c 127 g 220 t

ORIGIN

Query Match 4.1%; Score 20; DB 10; Length 673;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 224 ATCAGCAGATAAATTAAG 243
|||||
Db 343 ATCAGCAGATAAATTAAG 324

RESULT 4
AZ731780/c 689 bp DNA linear GSS 25-JAN-2001
LOCUS RPCI-24-128C18.TJ RPCI-24 Mus musculus genomic clone RPCI-24-128C18
DEFINITION
; DNA sequence.
ACCESSION AZ731780
VERSION AZ731780.1 GI:12494163
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 689)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akirel,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24 unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 128 row: C column: 18
Seq primer: SP6
Class: BAC ends.

FEATURES
source
1..689
location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-128C18"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: PTARBAC1; Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the PTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 166 a 163 c 155 g 205 t

ORIGIN

Query Match 4.1%; Score 20; DB 12; Length 689;

Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 409 GCAGGGCAGACGACGAT 428
|||||
Db 543 GCAGGGCAGACGACGAT 524

RESULT 5
CNS06GLH/c 988 bp DNA linear GSS 17-JUN-2001
LOCUS T3 end of clone AS0AA005E09 of library AS0AA from strain CLIB 533
DEFINITION
of Saccharomyces bayanus, genomic survey sequence.
ACCESSION AL397851
VERSION AL397851.1 GI:12150837
KEYWORDS GSS.
SOURCE Saccharomyces bayanus.
ORGANISM Saccharomyces bayanus.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
1 (bases 1 to 988)
Bon,E., Neuvéglise,C., Casaregola,S., Artiguenave,F., Wincker,P., Aigle,M. and Durrens,P.
Genomic exploration of the hemiascomycetous yeasts: 5. Saccharomyces bayanus var. uvarum
FEBS Lett. 487 (1), 37-41 (2000)

REFERENCE
2 (bases 1 to 988)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brotlier,P., Casaregola,S., deMontigny,J., Dujon,B., Durrens,P., Lepingle,A., Lorente,B., Malpertuy,A., Neuvéglise,C., Ozier-Kalogiropoulos,O., Potter,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)

TITLE
JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

3 (bases 1 to 988)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seqrefgenoscope.cns.fr - Web: www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES
source
1..988
location/Qualifiers
/organism="Saccharomyces bayanus"
/strain="CLIB 533"
/variety="uvarum"
/db_xref="taxon:4931"
/clone="AS0AA005E09"
/clone_lib="AS0AA"
/note="end: T3"
/note="end: T3"
/note="end: T3"
/note="similar to Saccharomyces cerevisiae ORF YER075c [PP3 : protein tyrosine phosphatase]"
/evidence="not-experimental"

BASE COUNT 313 a 253 c 162 g 255 t

ORIGIN

Query Match 4.1%; Score 20; DB 12; Length 988;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 251 AGAGTTTGGAAAGCTAAAA 270
 Db 622 AGAGTTTGGAAAGCTAAAA 603

RESULT 6
 LOCUS A0003546
 DEFINITION A0003546 Selected chromosome 21 cDNA library Homo sapiens EST 04-DEC-1997
 ACCESSION A0003546
 VERSION A0003546
 KEYWORDS EST.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE Szulzewsky, I., Hunt, E., Nguyen, M., Korn, B., Roehrdanz, B., Lehrach, H. and Yaspo, M.L.H.
 JOURNAL An integrated transcript map for the whole human chromosome 21
 COMMENT Unpublished (1997)
 CONTACT: Yaspo, M.-L.
 Max Planck Institut fuer Molekulare Genetik
 Ihnestrasse 73, D14195 Berlin-Dahlem, Germany.

FEATURES
 source
 Location/Qualifiers
 1..376
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="21q"
 /clone="Mpi112-9K8"
 /clone_lib="Selected chromosome 21 cDNA library"

BASE COUNT 106 a 81 c 81 g 103 t 5 others
 ORIGIN

Query Match
 Best Local Similarity 3.9%; Score 19; DB 9; Length 376;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 406 ATTGCAGGCGCAGAGCANG 424
 Db 37 ATTGCAGGCGCAGAGCANG 55

RESULT 7
 LOCUS T59966
 DEFINITION y67608.r1 Stragene ovary (4937217) Homo sapiens linear EST 09-FEB-1995
 IMAGE:76262 5' similar to SP:RCPD_MOUSE P80315 T-COMPLEX PROTEIN 1,
 ACCESSION T59966
 VERSION T59966.1 GI:661803
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, S., Morris, M., Parsons, J., Prange, C., Rifkin, B., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E., Underwood, K., Woldman, P., Waterston, R., Wilson, R. and Marra, M.
 JOURNAL Generation and analysis of 280,000 human expressed sequence tags
 MEDLINE Genome Res. 6 (9), 807-828 (1996)
 COMMENT 97044478

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

FEATURES
 source
 Location/Qualifiers
 1..380
 /organism="Homo sapiens"
 /db_xref="GDB:505991"
 /db_xref="taxon:9606"
 /clone="IMAGE:76262"
 /clone_lib="Stragene ovary (4937217)"
 /sex="Female"
 /dev_stage="49 year old"
 /lab_host="SOLR cells (kanamycin resistant)"
 /note="Organ: ovary; Vector: Bluescript SK; Site: 1: EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Total ovary tissue, normal, caucasian. Average insert size: 0.8 kb. Uni-ZAP XR Vector. -5' adaptor insert GAATTCGCGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."
 BASE COUNT 104 a 68 c 93 g 111 t 4 others
 ORIGIN

Query Match
 Best Local Similarity 3.9%; Score 19; DB 10; Length 380;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 247 AACACAGTTTGGAAAGG 265
 Db 283 AACACAGTTTGGAAAGG 301

RESULT 8
 LOCUS AW151916
 DEFINITION x172906.x1 NCI-CGAP-Gas4 Homo sapiens cDNA clone IMAGE:2623642 3'
 LOCUS AW151916
 ACCESSION AW151916
 VERSION AW151916.1 GI:6199901
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 COMMENT Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 CDNA Sequencing by: Washington University Genome Sequencing Center
 found through the I.M.A.G.E. Consortium/LNL at:
 www-bio.lnl.gov/bbcp/image/image.html

FEATURES
 source
 Location/Qualifiers
 1..436
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2623642"
 /clone_lib="NCI-CGAP-Gas4"

Possible reversed clone: polyA not found
 Seq primer: -40up from Glibco
 High quality sequence stop: 412.
 Location/Qualifiers
 1..436
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2623642"
 /clone_lib="NCI-CGAP-Gas4"


```

/tissue_type="poorly differentiated adenocarcinoma with
signed ring cell features"
/lab_host="DH10B"
/note="Organ: Stomach; Vector: pCMV-SPORT6; Site_1: Salt,
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"

BASE COUNT      136 a      98 c      86 g      116 t
ORIGIN

Query Match      3.9%; Score 19; DB 9; Length 436;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      299 CAGTACAGGCAATGCTTA 317
Db      256 CAGTACAGGCAATGCTTA 238

RESULT 9
AZ952777      461 bp      DNA      linear      GSS 27-APR-2001
LOCUS      2M0217C15R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
DEFINITION
ACCESSION      AZ952777
VERSION      AZ952777.1 GI:13824004
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
1 (bases 1 to 461)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weis,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0217 row: C column: 15
Seq primer: CACACAGGAACAGCTGTGACC
Class: plasmid ends
High quality sequence stop: 461.
Location/Qualifiers
1. 461
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0217C15"
/clone_1lb="Mouse 10kb plasmid UUGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: pMD29v; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative

```

```

of pMD42 (g14732114|p|AF129072.1), a copy-number
inducible derivative of plasmid RL. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT      123 a      91 c      121 g      126 t
ORIGIN

Query Match      3.9%; Score 19; DB 12; Length 461;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      170 TTGGTGCTGCTTTTGGG 188
Db      413 TTGGTGCTGCTTTTGGG 431

RESULT 10
A1965711      545 bp      mRNA      linear      EST 30-NOV-2001
LOCUS      sc77604.y1 Gm-c1018 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION      Gm-c1018-919 5' similar to TR:080558 080558 T22J18.21 PROTEIN.;
ACCESSION      A1965711
VERSION      A1965711.1 GI:5760420
KEYWORDS      EST.
SOURCE      soybean.
ORGANISM      Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 545)
Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,D., Coryell,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers
,T., Person,B., Swaller,T., Gibbons,M., Pepe,D., Harvey,N., Schuk
,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Possible reversed clone: similarity on wrong strand This clone is
available through: Resgen, Invitrogen Corp. 2130 South Memorial
Parkway Huntsville, AL 35801 For further information call: (800
)-533-4363 or contact via email: ccu@resgen.com
High quality sequence stop: 421.
Location/Qualifiers
1. 545
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1018-919"
/clone_1lb="Gm-c1018"
/tissue_type="leaves of greenhouse grown plants"
/dev_stage="2-3 weeks old"
/lab_host="DH10B (Gibco BRL)"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from 2-3 week old greenhouse grown plants. The cDNA library
was prepared using the Stratagene pluscript II XR
library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adaptors
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally

```

cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelidg.

BASE COUNT 147 a 115 c 120 g 163 t
 Query Match 3.9%; Score 19; DB 9; Length 545;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 181 GTTTGGGTGATTAATG 199
 ||||||||||||||||
 Db 306 GTTTGGGTGATTAATG 324

RESULT 11
 AZ887587
 LOCUS RPCI-24-166M15.TJ RPCI-24 Mus musculus genomic clone RPCI-24-166M15
 DEFINITION
 ' DNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 550)
 Zhao,S., Nierman,W., Malek,J., Shatsman,S., Aklnet,B., Levins,M.,
 Tsegaye,G., Geer,K., Krol,M., Shwartsbeyn,A., Gebregorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Other.GSSS: RPCI-24-166M15.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end plates: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
 Plate: 166 row: M column: 15
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 Source Location/Qualifiers
 1..550
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-166M15"
 /clone_1ib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARAPAC1, Site_1: BamHI, Site_2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTARAPAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 126 a 155 c 123 g 146 t
 ORIGIN
 Query Match 3.9%; Score 19; DB 12; Length 550;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 170 TTGGTGGTCTGTTTGGG 188
 ||||||||||||||||

Db 213 TTGGTGGTCTGTTTGGG 231
 ||||||||||||||||

RESULT 12
 BM091630
 LOCUS BM091630
 DEFINITION
 BM091630.y1 Gm-c1086 glycine max cDNA clone GENOME SYSTEMS CLONE
 ID: Gm-c1086-203 5' similar to TR:09T0C8 09T0C8 PUTATIVE
 PHOSPHORIBOSYLAMINTRANSLATE TRANSFERASE. ; mRNA sequence.
 BM091630
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 soybean.
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 561)
 Shoemaker,R., Keim,P., Vodkin,L., Erpelidg,J., Corryell,V., Khanna
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Rheising,B., Allen,M., Bowers
 Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk
 R., Ralter,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann
 R., Waterson,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

This clone is available through: Resgen, Invitrogen Corp, 2130
 South Memorial Parkway Huntville, AL 35801 For further information
 call: (800)-533-4363 or contact via email: cvu@resgen.com
 High quality sequence stop: 421.

FEATURES
 Source Location/Qualifiers
 1..561
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1086-203"
 /clone_1ib="Gm-c1086"
 /tissue_type="Young seeds (Williams 82)"
 /lab_host="DH10B"
 /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from mRNA isolated from very young seeds (less than 20mgs). The library was prepared using the Stratagene pBluescript II SK (+) library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with an xho I restriction site. Eco RI adaptors were ligated to the blunt-ended cDNA fragments followed by xho I digestion. The cDNA insert is protected from xho I digestion via methylation during first strand cDNA synthesis. The cDNA fragments were directionally cloned into the Eco RI-Xho I restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cell. The library was constructed by Anu Khanna (Alia Vodkin lab, University of Illinois)."

BASE COUNT 126 a 136 c 134 g 165 t
 ORIGIN
 Query Match 3.9%; Score 19; DB 10; Length 561;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 181 GTTTGGGTGATTAATG 199
 ||||||||||||||||
 Db 4 GTTTGGGTGATTAATG 22

RESULT 13
BM082634/c 582 bp mRNA linear EST 14-NOV-2001
LOCUS f423j09.y1 Campbell zebrafish ovary danio rerio CDNA clone 5306345
DEFINITION 5' similar to TR:Q9Y6R2 Q9Y6R2 EMBRYONIC LUNG PROTEIN. ; mRNA
sequence.
ACCESSION BM082634 GI:16929564
VERSION BM082634.1
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
REFERENCE
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
TITLE Mashu zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.wustl.edu
CDNA library constructed by library constructed by Invitrogen and
donated by R. Campbell (Marine Biology Laboratory, Woods Hole, MA).
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: the I.M.A.G.E. Consortium/LNL, send email to:
info@image.llnl.gov
High quality sequence stop: 440.
Location/Qualifiers
FEATURES
source 1..582
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_id="5306345"
/clone_lib="Campbell zebrafish ovary"
/sex="female"
/dev_stage="4-5 months, 1 year and 2 years"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary(pooled); Vector: pCMV-SPORT6; Site:1;
NCBI: Site_2: Salt; Whole ovaries collected from zebrafish
aged 4-5 months, 1 year and 2 years. Oligo-dt primed,
directionally cloned. Average insert size 2 kb. Library
constructed by Invitrogen and donated by R. Campbell
(Marine Biology Laboratory, Woods Hole, MA)."
BASE COUNT 166 a 145 c 137 g 133 t 1 others
ORIGIN
Query Match 3.9%; Score 19; DB 10; Length 582;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 240 AAACCTAACGAGTTTG 258
|||||
Db 207 AAACCTAACGAGTTTG 189
RESULT 14
BM083068 584 bp mRNA linear EST 14-NOV-2001
LOCUS f428h07.y1 Campbell zebrafish ovary danio rerio CDNA clone 5306748
DEFINITION 5' similar to TR:Q9Y6R2 Q9Y6R2 EMBRYONIC LUNG PROTEIN. ; mRNA
sequence.
ACCESSION BM083068 GI:16929998
VERSION BM083068.1
KEYWORDS EST.

SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
REFERENCE
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
TITLE Mashu zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Other_ESTs: f428h07.x1
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.wustl.edu
CDNA library constructed by library constructed by Invitrogen and
donated by R. Campbell (Marine Biology Laboratory, Woods Hole, MA).
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: the I.M.A.G.E. Consortium/LNL, send email to:
info@image.llnl.gov
High quality sequence stop: 442.
Location/Qualifiers
FEATURES
source 1..584
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_id="5306748"
/clone_lib="Campbell zebrafish ovary"
/sex="female"
/dev_stage="4-5 months, 1 year and 2 years"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary(pooled); Vector: pCMV-SPORT6; Site:1;
NCBI: Site_2: Salt; Whole ovaries collected from zebrafish
aged 4-5 months, 1 year and 2 years. Oligo-dt primed,
directionally cloned. Average insert size 2 kb. Library
constructed by Invitrogen and donated by R. Campbell
(Marine Biology Laboratory, Woods Hole, MA)."
BASE COUNT 170 a 149 c 136 g 129 t
ORIGIN
Query Match 3.9%; Score 19; DB 10; Length 584;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 240 AAACCTAACGAGTTTG 258
|||||
Db 209 AAACCTAACGAGTTTG 191
RESULT 15
BI474188 585 bp mRNA linear EST 27-AUG-2001
LOCUS fp36b02.y3 zebrafish gridless kidney danio rerio CDNA clone 4745043
DEFINITION 5' similar to TR:Q9Y6R2 Q9Y6R2 EMBRYONIC LUNG PROTEIN. ; mRNA
sequence.
ACCESSION BI474188
VERSION BI474188.1 GI:15303352
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
REFERENCE
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.

TITLE
JOURNAL
COMMENT

Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
 and Wilson, R.
 Mashu zebrafish EST Project 1998
 Unpublished (1998)
 Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrafish@watson.wustl.edu
 cDNA Library Preparation: Leonard Ira Zon DNA Sequencing by:
 Washington University Genome Sequencing Center Clone Distribution:
 Genome Systems, St. Louis, Missouri (web address:
 www.genomesystems.com) (email contact: info@genomesystems.com)
 Research Genetics, Huntsville, Alabama (web address: www.resgen.com)
) (email contact: info@resgen.com) and
 RessourcenzentrumPrimatDatenbank, Berlin, Germany (web address:
 www.rzpd.de)
 High quality sequence stop: 320.

FEATURES
source

Location/Qualifiers
 1. 585
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone="4745043"
 /clone_lib="zebrafish gridded kidney"
 /sex="mixed"
 /tissue_type="kidney pooled from 300 wild type adults"
 /lab_host="XLOLR"
 /note="Organ: kidney; Vector: pBK-CMV; Site 1: EcorI;
 Site 2: XhoI; Oligo dt cDNA library constructed from mRNA
 pooled from pooled kidney tissue from 300 adult
 zebrafish."
 BASE COUNT 161 a 140 c 136 g 146 t 2 others
 ORIGIN

Query Match 3.9%; Score 19; DB 10; Length 585;
 Best Local Similarity 100.0%; Prid. No. 27;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 240 AAAGCTAAACGAGCTTG 258
 ||||||||||||||||
 DB 291 AAAGCTAAACGAGCTTG 273

Search completed: October 27, 2002, 22:36:36
 Job time : 978.679 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 20:19:49 : Search time 36.0933 Seconds
(without alignments)
3307.484 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486
Sequence: 1 atgaacagagatgttgc.....aagtcattcacacagaaaa 486

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- Issued_Patents_NA:*
- 1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
 - 2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
 - 3: /cgn2_6/ptodata/2/1na/6A.COMB.seq:*
 - 4: /cgn2_6/ptodata/2/1na/6B.COMB.seq:*
 - 5: /cgn2_6/ptodata/2/1na/PCFUS.COMB.seq:*
 - 6: /cgn2_6/ptodata/2/1na/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	18	3.7	1331	3 US-08-985-908-23	Sequence 23, Appl
2	18	3.7	3776	1 US-08-162-809-7	Sequence 7, Appl
3	17	3.5	2165	2 US-08-669-721-2	Sequence 2, Appl
4	17	3.5	2165	4 US-09-189-344-2	Sequence 2, Appl
5	17	3.5	2340	3 US-08-742-877-3	Sequence 3, Appl
6	17	3.5	2775	4 US-09-053-871A-22	Sequence 22, Appl
7	17	3.5	2802	3 US-08-742-877-1	Sequence 1, Appl
8	17	3.5	51952	3 US-08-947-823-1	Sequence 1, Appl
9	17	3.5	4403765	4 US-09-103-840A-2	Sequence 2, Appl
10	16	3.3	1632	1 US-07-959-941-1	Sequence 1, Appl
11	16	3.3	1632	1 US-08-259-924-1	Sequence 1, Appl
12	16	3.3	2251	4 US-08-991-677-11	Sequence 11, Appl
13	16	3.3	3429	1 US-08-097-997A-10	Sequence 10, Appl
14	16	3.3	3429	3 US-08-665-574C-10	Sequence 10, Appl
15	16	3.3	3429	4 US-08-946-994-10	Sequence 10, Appl
16	16	3.3	4234	1 US-08-446-038B-1	Sequence 1, Appl
17	16	3.3	4234	1 US-08-446-010B-1	Sequence 1, Appl
18	16	3.3	4234	1 US-08-805-445-1	Sequence 1, Appl
19	16	3.3	4234	2 US-08-064-067D-1	Sequence 1, Appl
20	16	3.3	4234	2 US-09-066-208-1	Sequence 1, Appl
21	16	3.3	4868	5 US-08-139-937-12	Sequence 12, Appl
22	16	3.3	4868	5 PCT-US93-11310-12	Sequence 12, Appl
23	16	3.3	5703	4 US-08-280-590A-36	Sequence 36, Appl
24	16	3.3	8789	1 US-08-328-234-5	Sequence 5, Appl
25	16	3.3	10136	1 US-08-353-700-2	Sequence 2, Appl
26	16	3.3	10136	5 PCT-US95-16216-2	Sequence 2, Appl
27	16	3.3	24417	2 US-08-846-762-1	Sequence 1, Appl

c 28	15	3.1	59	4 US-09-037-990B-64	Sequence 64, Appl
c 29	15	3.1	99	1 US-08-248-474-77	Sequence 77, Appl
c 30	15	3.1	98	3 US-08-756-849-77	Sequence 77, Appl
c 31	15	3.1	162	4 US-09-037-990B-36	Sequence 36, Appl
c 32	15	3.1	177	1 US-08-065-845-9	Sequence 9, Appl
c 33	15	3.1	177	1 US-08-429-523-9	Sequence 9, Appl
c 34	15	3.1	177	1 US-08-429-532-9	Sequence 9, Appl
c 35	15	3.1	177	1 US-08-429-532-9	Sequence 9, Appl
c 36	15	3.1	177	1 US-08-429-532-9	Sequence 9, Appl
c 37	15	3.1	177	1 US-08-429-532-9	Sequence 9, Appl
c 38	15	3.1	249	4 US-08-269-136B-9	Sequence 9, Appl
c 39	15	3.1	249	3 US-08-248-474-4	Sequence 4, Appl
c 40	15	3.1	354	2 US-08-756-849-4	Sequence 4, Appl
c 41	15	3.1	379	4 US-08-676-279-38	Sequence 38, Appl
c 42	15	3.1	422	4 US-08-838-151A-58	Sequence 58, Appl
c 43	15	3.1	494	4 US-09-040-984-54	Sequence 12, Appl
c 44	15	3.1	494	4 US-09-123-912-54	Sequence 54, Appl
c 45	15	3.1	725	3 US-09-197-801-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-985-908-23
Sequence 23, Application US/08985908
Patent No. 6004773
GENERAL INFORMATION:
APPLICANT: MASAYUKI ARAKI, MASAKAZU SUGIMOTO, YASUHIKO YOSHINARA, AND TSUYOSHI
TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,908
FILING DATE: 05-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-325659
FILING DATE: 05-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1331 bases
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORGANISM: Escherichia coli
STRAIN: JM109
FEATURE:
NAME/KEY: CDS
LOCATION: 10..1197
US-08-985-908-23
Query Match 3.7%; Score 18; DB 3; Length 1331;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 GCTGTTGGTAAAGTAG 149
|||||
Db 252 GCTGTTGGTAAAGTAG 269

RESULT 2

US-08-162-809-7
Sequence 7, Application US/08162809
Patent No. 5457048
GENERAL INFORMATION:
APPLICANT: Pasquale, Elena B.
APPLICANT: Sajjadi, Rereyoun G.
TITLE OF INVENTION: NOVEL EPH-RELATED TYROSINE KINASES,
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States of America
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162,809
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9503
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3776 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 290..3208
US-08-162-809-7

Query Match 3.7%; Score 18; DB 1; Length 3776;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 GGCATGTGGTGGCTGT 182
|||||
Db 271 GGCATGTGGTGGCTGT 288

RESULT 3

US-08-669-721-2/c
Sequence 2, Application US/08669721
Patent No. 5834236
GENERAL INFORMATION:
APPLICANT: Lamb et al., Christopher J.
TITLE OF INVENTION: NOVEL TRANSCRIPTION ENHANCER ELEMENT AND
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA

COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/669,721

FILING DATE: 27-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ellison, Eldora L.

REGISTRATION NUMBER: 39,967
REFERENCE/DOCKET NUMBER: 07251/014001

TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 2165 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS
LOCATION: 61..1698

US-08-669-721-2

Query Match

Best Local Similarity 3.5%; Score 17; DB 2; Length 2165;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 ACTGGGCTGTGCTG 116
|||||
Db 673 ACTGGGCTGTGCTG 657

RESULT 4

US-09-189-344-2/c
Sequence 2, Application US/09189344
Patent No. 6191258
GENERAL INFORMATION:
APPLICANT: Lamb et al., Christopher J.
TITLE OF INVENTION: NOVEL TRANSCRIPTION ENHANCER ELEMENT AND
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/189,344
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/669,721
FILING DATE: 27-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ellison, Eldora L.
REGISTRATION NUMBER: 39,967
REFERENCE/DOCKET NUMBER: 07251/014001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2165 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 61..1698
US-09-189-344-2

Query Match 3.5%; Score 17; DB 4; Length 2165;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 ACTGGGCTGTTGTCG 116
|||||
DB 673 ACTGGGCTGTTGTCG 657

RESULT 5

US-08-742-877-3/C
Sequence 3, Application US/08742877
Patent No. 6046380
GENERAL INFORMATION:
APPLICANT: CLARK, Anthony J.
TITLE OF INVENTION: DNA SEQUENCES
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,877
FILING DATE: 01-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9408717.8
FILING DATE: 03-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: FLESHNER, RAZ E.
REGISTRATION NUMBER: 34,331
REFERENCE/DOCKET NUMBER: 0623.0470001/REF
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2340 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: mRNA
US-08-742-877-3

Query Match 3.5%; Score 17; DB 3; Length 2340;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 255 TTGGAAGGTAAG 271
|||||
DB 1495 TTGGAAGGTAAG 1479

RESULT 6

US-09-053-871A-22/C
Sequence 22, Application US/09053871A
Patent No. 6315995
GENERAL INFORMATION:
APPLICANT: Pinsky, David J.
APPLICANT: Stern, David
APPLICANT: Rose, Eric
APPLICANT: Solomon, Robert A.
APPLICANT: Schmidt, Ann Marie
TITLE OF INVENTION: METHODS FOR TREATING AN ISCHEMIC DISORDER AND IMPROVING
FILE REFERENCE: 51917-B
CURRENT APPLICATION NUMBER: US/09/053,871A
CURRENT FILING DATE: 1998-04-01
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 22
LENGTH: 2775
TYPE: DNA
ORGANISM: Homo Sapien
US-09-053-871A-22

Query Match 3.5%; Score 17; DB 4; Length 2775;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 255 TTGGAAGGTAAG 271
|||||
DB 1928 TTGGAAGGTAAG 1912

RESULT 7

US-08-742-877-1/C
Sequence 1, Application US/08742877
Patent No. 6046380
GENERAL INFORMATION:
APPLICANT: CLARK, Anthony J.
TITLE OF INVENTION: DNA SEQUENCES
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,877
FILING DATE: 01-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9408717.8
FILING DATE: 03-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: FLESHNER, RAZ E.
REGISTRATION NUMBER: 34,331
REFERENCE/DOCKET NUMBER: 0623.0470001/REF
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2802 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: mRNA
FEATURE:

NAME/KEY: CDS
LOCATION: 30..1412
FEATURE:
NAME/KEY: sig-peptide
LOCATION: 30..167
FEATURE:
NAME/KEY: mat-peptide
LOCATION: 168..1412
US-08-742-877-1

Query Match 3.5%; Score 17; DB 3; Length 2802;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 255 TTGGGAAAGGTAAAG 271
DB 1957 TTGGGAAAGGTAAAG 1941

RESULT 8
US-08-947-823-1
Sequence 1, Application US/08947823
Patent No. 6114605
GENERAL INFORMATION:
APPLICANT: Williamson, Valerie M.
APPLICANT: Kaloshian, Isouh
APPLICANT: Yaghoobi, Jafar
APPLICANT: Bodeau, John
APPLICANT: Milligan, Stephen
TITLE OF INVENTION: Procedures and Materials for Conferring
TITLE OF INVENTION: Pest Resistance in Plants
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/947,823
FILING DATE: 09-OCT-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/18802
FILING DATE: 09-OCT-1997
APPLICATION NUMBER: US 60/028,191
FILING DATE: 10-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REFERENCE/DOCKET NUMBER: 023070-0702100S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 51952 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-947-823-1

Query Match 3.5%; Score 17; DB 3; Length 51952;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 470 TCATTCAACAGAAAA 486
DB 49389 TCATTCAACAGAAAA 49405

RESULT 9
US-09-103-840A-2/C
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match 3.5%; Score 17; DB 4; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 115 GCGCGTGTCTGCGCA 131
DB 4088325 GCGCGTGTCTGCGCA 4088309

RESULT 10
US-07-959-941-1
Sequence 1, Application US/07959941
Patent No. 5364781
GENERAL INFORMATION:
APPLICANT: HUTCHINSON, Charles R.
APPLICANT: WADDURI, Krishna M.
APPLICANT: TORI, Francesca
APPLICANT: COLOMBO, Anna L.
TITLE OF INVENTION: PROCESS FOR PREPARING DAUNORUBICIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nikaido, Marmelstein, Murray & Oram
STREET: 655 Fifteenth Street N.W. Suite 330
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/959,941
FILING DATE: 19921009
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 793,873
FILING DATE: 18-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: Chin, Monica F.

REGISTRATION NUMBER: P-36,105
REFERENCE/DOCKET NUMBER: 1615-1816CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)638-5000
TELEFAX: (202)638-4810
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1632 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 204..1271
US-07-959-941-1

Query Match 3.3%; Score 16; DB 1; Length 1632;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 439 GCATGCCGCAACCGG 454
Db 1 GCATGCCGCAACCGG 16
|||||

RESULT 11
US-08-259-924-1
Sequence 1, Application US/08259924
Patent No. 5563064
GENERAL INFORMATION:
APPLICANT: HUTCHINSON, Charles R.
APPLICANT: MADURIT, Krishna M.
APPLICANT: TORRI, Francesca
APPLICANT: COLOMBO, Anna L.
TITLE OF INVENTION: PROCESS FOR PREPARING DAUNORUBICIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nikaido, Marmelstein, Murray & Oram
STREET: 655 Fifteenth Street N.W. Suite 330
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/259,924
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 793,873
FILING DATE: 18-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/959,941
FILING DATE: 09-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Kites, Monica C
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: 1615-4003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)638-5000
TELEFAX: (202)638-4810
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1632 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

FEATURE:
NAME/KEY: CDS
LOCATION: 204..1271
US-08-259-924-1

Query Match 3.3%; Score 16; DB 1; Length 1632;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 439 GCATGCCGCAACCGG 454
Db 1 GCATGCCGCAACCGG 16
|||||

RESULT 12
US-08-991-677-11
Sequence 11, Application US/08991677A
Patent No. 6252135
GENERAL INFORMATION:
APPLICANT: Chiang, Vincent L.
APPLICANT: Carraway, Daniel T.
APPLICANT: Smeltzer, Richard H.
TITLE OF INVENTION: Production of Syringyl Lignin in Gymnosperms
FILE REFERENCE: 50617
CURRENT APPLICATION NUMBER: US/08/991,677A
CURRENT FILING DATE: 1997-12-16
EARLIER APPLICATION NUMBER: US 60/033,381
EARLIER FILING DATE: 1996-12-16
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 11
LENGTH: 2251
TYPE: DNA
ORGANISM: Pinus taeda
US-08-991-677-11

Query Match 3.3%; Score 16; DB 4; Length 2251;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 234 TAAATAAGCTAAGC 249
Db 310 TAAATAAGCTAAGC 325
|||||

RESULT 13
US-08-097-997A-10/C
Sequence 10, Application US/08097997A
Patent No. 5728536
GENERAL INFORMATION:
APPLICANT: Ihle, James N.
APPLICANT: Silvenoinen, Olli
APPLICANT: Wituhn, Bruce A.
APPLICANT: Ouellet, Frederick W.
TITLE OF INVENTION: Jak Kinases and Regulation of Cytokine Signal
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/097,997A
FILING DATE: 29-JULY-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Fox, Samuel L.
REGISTRATION NUMBER: 30,353
REFERENCE/DOCKET NUMBER: 0656.0370000/SLF/GKT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2540
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 3429 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3426
US-08-097-997A-10

Query Match

Best Local Similarity 3.3%; Score 16; DB 1; Length 3429;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 CTGGCCAGCTGTTGG 140
DB 2063 CTGGCCAGCTGTTGG 2048

RESULT 14

US-08-665-574C-10/C
Sequence 10, Application US/08665574C
Patent No. 6136595

GENERAL INFORMATION:

APPLICANT: Ihle, James N.
APPLICANT: Silvenoinen, Olli
APPLICANT: Wiltuhh, Bruce A.
TITLE OF INVENTION: Jak Kinases and Regulation of Cytokine
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,574C
FILING DATE: 18-JUN-1996

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/282,012

FILING DATE: 29-JUL-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/097,997

FILING DATE: 29-JUL-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/118,968

FILING DATE: 09-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: Bugalsky, Lawrence B.

REGISTRATION NUMBER: 35,086

REFERENCE/DOCKET NUMBER: 0656.0370002/SLF/LBB

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 3429 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS

LOCATION: 1..3426

US-08-665-574C-10

Query Match

Best Local Similarity 3.3%; Score 16; DB 3; Length 3429;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 CTGGCCAGCTGTTGG 140
DB 2063 CTGGCCAGCTGTTGG 2048

RESULT 15

US-08-946-994-10/C
Sequence 10, Application US/08946994
Patent No. 6210654

GENERAL INFORMATION:

APPLICANT: Ihle, James N.
APPLICANT: Silvenoinen, Olli
APPLICANT: Wiltuhh, Bruce A.
APPLICANT: Quille, Frederick W.
TITLE OF INVENTION: Jak Kinases and Regulation of Cytokine Signal
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,994
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/665,574

FILING DATE: 18-JUN-1996

APPLICATION NUMBER: 08/282,012

FILING DATE: 29-JUL-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/097,997

FILING DATE: 29-JUL-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/118,968

FILING DATE: 09-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: Fox, Samuel L.

REGISTRATION NUMBER: 30,353

REFERENCE/DOCKET NUMBER: 0656.0370002/SLF/GKT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 3429 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS

LOCATION: 1..3426

US-08-946-994-10

Query Match 3.3%; Score 16; DB 4; Length 3429;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 125 CTGCCCAGCTGTTGG 140
|||||
Db 2063 CTGCCCAGCTGTTGG 2048

Search completed: October 27, 2002, 22:53:59
Job time : 942.093 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 17:46:54 : Search time 123.109 Seconds
(without alignments)
6777.910 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486
Sequence: 1 atgacagagagatgttgcga.....aagtcattcaacagaaaa 486

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N.Geneseq_032802.*

```
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	486	100.0	486	22 AAF86246	Ospa 17kd antigen
2	486	100.0	489	22 AAH79040	Piscirickettsia sa
3	23	4.7	33	22 AAF86249	Forward Ospa PCR p
4	22	4.5	40	22 AAF86250	Reverse Ospa PCR p
5	20	4.1	50000	21 AAA96363	Polymorphic repeat
6	19	3.9	1300	21 AAC48250	Arabidopsis thalia
7	19	3.9	1302	21 AAC47949	Arabidopsis thalia
8	18	3.7	401	22 AAS26580	Human cDNA encodin
9	18	3.7	932	21 AAF16080	Human prostate can

10	18	3.7	1293	8 AAN71108	Optimized Escheric
11	18	3.7	1331	19 AAV40259	Escherichia coli a
12	18	3.7	3659	8 AAN71109	phea arcf aspc ope
13	18	3.7	3776	16 AAO90655	eph-related prk Ce
14	18	3.7	9813	20 AAX13059	Enterococcus faeca
15	18	3.7	10432	23 ABL10082	Drosophila melanog
16	18	3.7	14250	23 ABL21060	Drosophila melanog
17	18	3.7	14759	22 AAS46273	DNA encoding novel
18	18	3.7	53552	22 AAS13655	Genomic DNA sequen
19	18	3.5	340	22 AAI90320	Human polynucleoti
20	17	3.5	387	22 AAI83097	Human polynucleoti
21	17	3.5	402	22 ABA26539	Probe #5005 for ge
22	17	3.5	402	22 AAK05021	Human brain expres
23	17	3.5	430	21 AAC43569	Zea mays DNA fragm
24	17	3.5	632	21 AAC44035	Zea mays DNA fragm
25	17	3.5	731	21 AAC36443	Pinus radiata tran
26	17	3.5	862	22 AAH03544	Human cDNA clone (
27	17	3.5	995	21 AAC34118	Arabidopsis thalia
28	17	3.5	1272	22 AAF54079	hFIX gene AE3' age
29	17	3.5	1272	22 AAF54080	hFIX gene AE3' age
30	17	3.5	1273	22 AAF54017	hFIX gene 3' UTR a
31	17	3.5	1273	22 AAF54073	hFIX gene AE3' age
32	17	3.5	1273	22 AAF54074	hFIX gene AE3' age
33	17	3.5	1273	22 AAF54075	hFIX gene AE3' age
34	17	3.5	1273	22 AAF54076	hFIX gene AE3' age
35	17	3.5	1275	22 AAF54077	hFIX gene AE3' age
36	17	3.5	1276	22 AAF54078	hFIX gene AE3' age
37	17	3.5	1285	22 AAF88311	C. officinalis cal
38	17	3.5	1367	23 AAS90536	DNA encoding novel
39	17	3.5	1521	22 AAH68197	C glutamicum codin
40	17	3.5	1633	21 AAC93386	Human secreted pro
41	17	3.5	1633	21 AAC93386	Human factor IX ge
42	17	3.5	1971	22 ABA55251	Human foetal liver
43	17	3.5	2003	22 AAS23063	DNA encoding novel
44	17	3.5	2165	19 AAV09781	Tobacco PABF cDNA.
45	17	3.5	2315	23 ABL24224	Drosophila melanog

ALIGNMENTS

RESULT 1	AAAF86246	standard: DNA; 486 BP.
ID	AAAF86246	
AC	AAAF86246	
DT	11-JUL-2001	(first entry)
DE	Ospa 17kd antigen gene.	
XX		
KW	Polkiltothermic fish: Piscirickettsia salmonis; rickettsial pathogen;	
KW	vacine; Ospa: salmonid rickettsial septicemia; rickettsial disease;	
KW	SRS; ds.	
OS	Piscirickettsia salmonis.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..486
FT		/*tag= a
FT		/partial
FT		/product= "Ospa"
FT		/note= "Genus specific 17kda antigen, the sequence does not include a stop codon"
CA2281913-A1.		
17-MAR-2001.		
17-SEP-1999;	99CA-2281913.	
17-SEP-1999;	99CA-2281913.	

PA (KAYM) KAY W W.
PA (BURI) BURIAN J.
XX (KUZ) KUZYSK M A.

PI Kay WM, Burian J, Kuzysk MA;

DR WPI: 2001-316844/34.
DR P-PSDB: AAB81126.

PT Method for protecting polkiothermic fish against salmonid rickettsial
PT vaccine containing the Ospa protein of Piscirickettsia salmonis
XX
PS Disclosure; Fig 2B; 35pp; English.

CC This invention relates to a method for the protection against infection
CC of a polkiothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkiothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
CC sequence represents DNA encoding the P. salmonis Ospa protein. An Ospa
CC protein with an N-terminal fusion partner is used in a vaccine to create
CC an anti-Ospa antibody response.

SO Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other:

Query Match Best Local Similarity 100.0%; Score 486; DB 22; Length 486;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGACAGAGATGTTGCAAGTAGTATGTAATTTATTCAGTGTGTTTACTGGC 60
DB 1 ATGACAGAGATGTTGCAAGTAGTATGTAATTTATTCAGTGTGTTTACTGGC 60
OY 61 TGTGCCAGAACTTACTGCTCAAGAGAGTGGAGCTGGGCTGTTGTTGGCGT 60
DB 61 TGTGCCAGAACTTACTGCTCAAGAGAGTGGAGCTGGGCTGTTGTTGGCGT 60
OY 121 GTTCTGCGCCAGCTGTTTGTGTAAGTAGTGTGCAATGTCATGTTGTTGGTGT 180
DB 121 GTTCTGCGCCAGCTGTTTGTGTAAGTAGTGTGCAATGTCATGTTGTTGGTGT 180
OY 181 GTTTGGGTTGATTAATGTTCTAAATCGTCAATGATGATGAGAGATTAATA 240
DB 181 GTTTGGGTTGATTAATGTTCTAAATCGTCAATGATGATGAGAGATTAATA 240
OY 241 AAGCTAAACCAAGTTGTAAGTAAGAGGAGGAGTGGGCTGTTGTTGGTGT 300
DB 241 AAGCTAAACCAAGTTGTAAGTAAGAGGAGGAGTGGGCTGTTGTTGGTGT 300
OY 301 GATACAGGCAATGTTAAAGTGTGAGCCAGTGGCTACTTACCAAGCTTAAATACCA 360
DB 301 GATACAGGCAATGTTAAAGTGTGAGCCAGTGGCTACTTACCAAGCTTAAATACCA 360
OY 361 GAGGTGCGCCAGCAATATGTTGAGCCAGTGGCTACTTACCAAGCTTAAATACCA 360
DB 361 GAGGTGCGCCAGCAATATGTTGAGCCAGTGGCTACTTACCAAGCTTAAATACCA 360
OY 421 CAAGAGATTTAGGAGCAATGTTGAGCAATTTGAGCAAAAGGAGTGTGAGGAGAG 420
DB 421 CAAGAGATTTAGGAGCAATGTTGAGCAATTTGAGCAAAAGGAGTGTGAGGAGAG 420
OY 481 GAAAAA 486
DB 481 GAAAAA 486

RESULT 2
AAH79040
ID AAH79040 standard; cDNA; 489 BP.

XX
AC AAH79040;
XX
DT 15-JAN-2002 (first entry)
XX
DE Piscirickettsia salmonis polynucleotide p10.6.
XX
KW Piscirickettsia salmonis; Piscirickettsiosis; salmonid rickettsial;
KW septicemia; SRS; surface antigen; vaccine; antibacterial; fish;
KW ATCC VR-1361; SS.
XX
OS Piscirickettsia salmonis.
XX
FH Key
FT CDS Location/Qualifiers
FT 1..489
XX /tag a
XX
PD WO200168865-A2.
XX
PF 20-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-GB01055.
XX
PR 11-MAR-2000; 2000GB-0005838.
PR 01-JUL-2000; 2000GB-0016080.
PR 01-JUL-2000; 2000GB-0016082.
PR 29-JUL-2000; 2000GB-0018599.
XX
PA (AQUA-) AQUA HEALTH EURO LTD.
XX
PI Simard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;
PI Burzio L;
XX WPI: 2001-639050/73.
XX P-PSDB; AAG78025.
XX
PT New nucleic acids encoding an amino acid sequence homologous to the
PT surface antigen present on Piscirickettsia salmonis are useful to
PT protect fish against Piscirickettsiosis -
XX
PS Claim 3; Fig 4; 25pp; English.

CC The invention relates to nucleic acid sequences and the encoded protein
CC of a least part of the surface antigen present on Piscirickettsia
CC salmonis for production of a vaccine with antibacterial activity to
CC protect fish against P. salmonis which causes Piscirickettsiosis, also
CC known as salmonid rickettsial septicemia.

SO Sequence 489 BP; 139 A; 79 C; 144 G; 127 T; 0 other:

Query Match Best Local Similarity 100.0%; Score 486; DB 22; Length 489;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGACAGAGATGTTGCAAGTAGTATGTAATTTATTCAGTGTGTTTACTGGC 60
DB 1 ATGACAGAGATGTTGCAAGTAGTATGTAATTTATTCAGTGTGTTTACTGGC 60
OY 61 TGTGCCAGAACTTACTGCTCAAGAGAGTGGAGCTGGGCTGTTGTTGGCGT 60
DB 61 TGTGCCAGAACTTACTGCTCAAGAGAGTGGAGCTGGGCTGTTGTTGGCGT 60
OY 121 GTTCTGCGCCAGCTGTTTGTGTAAGTAGTGTGCAATGTCATGTTGTTGGTGT 180
DB 121 GTTCTGCGCCAGCTGTTTGTGTAAGTAGTGTGCAATGTCATGTTGTTGGTGT 180
OY 181 GTTTGGGTTGATTAATGTTCTAAATCGTCAATGATGATGAGAGATTAATA 240
DB 181 GTTTGGGTTGATTAATGTTCTAAATCGTCAATGATGATGAGAGATTAATA 240
OY 241 AAGCTAAACCAAGTTGTAAGTAAGAGGAGGAGTGGGCTGTTGTTGGTGT 300
DB 241 AAGCTAAACCAAGTTGTAAGTAAGAGGAGGAGTGGGCTGTTGTTGGTGT 300

```

OY 301 GATACAGCAATAGTATAGTGTGAGCCAGTGCCTACTTACAGCGTTACAAATGAACAA 360
    |||||||
DB 301 GATACAGCAATAGTATAGTGTGAGCCAGTGCCTACTTACAGCGTTACAAATGAACAA 360
OY 361 GAGCGTCGCCAGCAATATTTGTCGAAATTTTCAGCAAAAAGCGGATGATTCAGGCGAGAG 420
    |||||||
DB 361 GAGCGTCGCCAGCAATATTTGTCGAAATTTTCAGCAAAAAGCGGATGATTCAGGCGAGAG 420
OY 421 CAAGAGATTTACGGCACTGCAATGCCGCAACCGATGCTGTGGCAAGTCATTTCACAA 480
    |||||||
DB 421 CAAGAGATTTACGGCACTGCAATGCCGCAACCGATGCTGTGGCAAGTCATTTCACAA 480
OY 481 GAAAAA 486
    ||||||
DB 481 GAAAAA 486

RESULT 3
AAf86249
ID AAF86249 standard; DNA: 33 BP.
AC AAF86249;
XX
XX 11-JUL-2001 (first entry)
DE
XX Forward Ospa PCR primer used in construction of pTYB1-17kD.
XX
XX Polkillothermic fish: Piscirickettsia salmonis; rickettsial pathogen;
KM vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
XX SRS; 17E2; fusion construct; PCR primer: ss.
XX
XX Piscirickettsia salmonis.
OS Synthetic.
XX
XX CA2281913-A1.
PN
XX 17-MAR-2001.
PD
XX 17-SEP-1999; 99CA-2281913.
PF
XX 17-SEP-1999; 99CA-2281913.
PR
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
XX (KUZV/) KUZV M A.
XX
XX Kay WW, Burian J, Kuzyk MA;
PI
XX WPI: 2001-316844/34.
DR
XX
XX Method for protecting polkillothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of Piscirickettsia salmonis -
XX
XX Example 5; Page 21; 35pp: English.
XX
XX This invention relates to a method for the protection against infection
CC of a polkillothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkillothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used to amplify the P. salmonis Ospa DNA
CC sequence. The primer is used in the construction of pTYB1-17kD, which is
CC used to express optimised Ospa antigen with a C-terminal fusion partner.
CC The Ospa fusion protein is used in a vaccine to create an anti-Ospa
CC antibody response.
XX
XX Sequence 33 BP; 13 A; 3 C; 11 G; 6 T; 0 other;
SO

```

```

Query Match 4.7%; Score 23; DB 22; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGACAGAGAGATTTGCAAGG 23
    |||||||
DB 11 ATGACAGAGAGATTTGCAAGG 33

RESULT 4
AAf86250/c
ID AAF86250 standard; DNA: 40 BP.
AC AAF86250;
XX
XX 11-JUL-2001 (first entry)
DE
XX Reverse Ospa PCR primer used in construction of pTYB1-17kD.
XX
XX Polkillothermic fish: Piscirickettsia salmonis; rickettsial pathogen;
KM vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
XX SRS; 17E2; fusion construct; PCR primer: ss.
XX
XX Piscirickettsia salmonis.
OS Synthetic.
XX
XX CA2281913-A1.
PN
XX 17-MAR-2001.
PD
XX 17-SEP-1999; 99CA-2281913.
PF
XX 17-SEP-1999; 99CA-2281913.
PR
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
XX (KUZV/) KUZV M A.
XX
XX Kay WW, Burian J, Kuzyk MA;
PI
XX WPI: 2001-316844/34.
DR
XX
XX Method for protecting polkillothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of Piscirickettsia salmonis -
XX
XX Example 5; Page 21; 35pp: English.
XX
XX This invention relates to a method for the protection against infection
CC of a polkillothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkillothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used to amplify the P. salmonis Ospa DNA
CC sequence. The primer is used in the construction of pTYB1-17kD, which is
CC used to express optimised Ospa antigen with a C-terminal fusion partner.
CC The Ospa fusion protein is used in a vaccine to create an anti-Ospa
CC antibody response.
XX
XX Sequence 40 BP; 8 A; 10 C; 7 G; 15 T; 0 other;
SO
Query Match 4.5%; Score 22; DB 22; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 465 GCAAGTCATTTCAACGAAAAA 486
    |||||||
DB 40 GCAAGTCATTTCAACGAAAAA 19

```

RESULT 5
 ID AAA96363 standard; DNA; 50000 BP.
 CC AAA96363;
 AC 08-FEB-2001 (first entry)
 DE Polymorphic repeat microsatellite sequences present in the CTLA4 locus.
 KW Autoimmune disease; polymorphic microsatellite repeat; PMR; CD28 gene;
 KW ICOS gene; CTLA4 gene; costimulatory receptor gene locus; CGRL; lupus;
 KW insulin-dependent diabetes mellitus; IDDM; Addison's disease; leprosy;
 KW thyroiditis; postpartum hypothyroidism; myasthenia gravis; thymoma;
 KW Hashimoto's disease; coeliac disease; ss.
 OS Homo sapiens.
 FH Key
 FT satellite
 FT Location/Qualifiers
 FT 5722..5746
 FT /tag= a
 FT /note= "sara41/42 microsatellite repeat"
 FT 6530..6597
 FT /tag= b
 FT /note= "sara43/44 microsatellite repeat"
 FT 19911..19956
 FT /tag= c
 FT /note= "pw210/211 microsatellite repeat"
 FT 23904..23957
 FT /tag= d
 FT /note= "sara45/46 microsatellite repeat"
 FT 27689..27780
 FT /tag= e
 FT /note= "sara17/18 microsatellite repeat"
 FT 30766..30801
 FT /tag= f
 FT /note= "sara19/20 microsatellite repeat"
 PN W0200056856-A2.
 XX 28-SEP-2000.
 PD 24-MAR-2000; 2000MO-US07938.
 PF 25-MAR-1999; 99US-0126215.
 PR (GEM) GENETICS INST INC.
 PA Ling V, Wu P, Gray GS;
 PI WPI; 2000-628257/60.
 DR Determining predisposition of humans to develop autoimmune disease
 XX Involves detecting polymorphic microsatellite repeat sequence within
 PT human costimulatory receptor gene locus
 PS Claim 2; Page 67-82; 160pp; English.
 CC Two human bacterial artificial chromosome (BAC) clones that included
 CC and flanked the human CTLA-4 locus were cloned and sequenced. The
 CC sequence data was assembled into a contiguous sequence that is presented
 CC in AAA96363-68. AAA96363-64 comprise BAC clone 22700, and AAA96365-68
 CC comprise BAC clone 22608. The sequences contain polymorphic
 CC microsatellite repeat (PMR) sequences. The specification describes a
 CC method for determining the predisposition of a human subject to develop
 CC autoimmune disease. The method comprises detecting a PMR sequence in the
 CC CD28, ICOS gene or CTLA4 gene of the human costimulatory receptor gene
 CC locus (hCGR). PMR sequences vary in length among individuals and can
 CC be amplified to generate products that differ in size. These products can
 CC then be detected by rapid and convenient high resolution processes. The
 CC method is useful for determining the predisposition of insulin-dependent
 CC diabetes mellitus (IDDM), Addison's disease, Graves disease, autoimmune

CC hypothyroidism, myasthenia gravis, thymoma, lupus, thyroiditis,
 CC postpartum thyroiditis, rheumatoid arthritis, Hashimoto's disease,
 CC coeliac disease and leprosy. PMR sequences within hCGR are useful as
 CC markers in a variety of assays and in the field of forensic medicine,
 CC disease diagnosis and human genome mapping.
 XX Sequence 50000 BP; 14521 A; 9882 C; 11191 G; 14406 T; 0 other;
 SO
 Query Match 4.1%; Score 20; DB 21; Length 50000;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 251 AGAGTTTGGAAGGTAATAA 270
 Db 3317 AGAGTTTGGAAGGTAATAA 3336
 RESULT 6
 AAC48250/C
 ID AAC48250 standard; DNA; 1300 BP.
 CC AAC48250;
 AC 18-OCT-2000 (first entry)
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 56805.
 XX Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 OS Arabidopsis thaliana.
 PN EP1033405-A2.
 XX 06-SEP-2000.
 PD 25-FEB-2000; 2000EP-0301439.
 PF 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 16-APR-1999; 99US-0128714.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 11-MAY-1999; 99US-0132487.
 PR 14-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 19-MAY-1999; 99US-0134376.
 PR 20-MAY-1999; 99US-0134941.
 PR 21-MAY-1999; 99US-0135124.
 PR 24-MAY-1999; 99US-0135353.
 PR 25-MAY-1999; 99US-0135629.
 PR 27-MAY-1999; 99US-0136021.
 PR 99US-0136392.

PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139809.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144652.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145112.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0165596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

Best Local Similarity 3.9%; Score 19; DB 21; Length 1300;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;DY 415 CAGACGACGACATTACG 433
DB 121 CAGACGACGACATTACG 103
|||||

RESULT 7

AAC47949/C

ID AAC47949 standard; DNA; 1302 BP.

AC AAC47949;

XX 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 55704.

KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 990S-0121825.

PR 09-MAR-1999; 990S-0121810.

PR 23-MAR-1999; 990S-0123548.

PR 25-MAR-1999; 990S-0125788.

PR 29-MAR-1999; 990S-0126264.

PR 01-APR-1999; 990S-0126785.

PR 08-APR-1999; 990S-0127462.

PR 16-APR-1999; 990S-0128714.

PR 19-APR-1999; 990S-0128845.

PR 21-APR-1999; 990S-0130077.

PR 23-APR-1999; 990S-0130443.

PR 28-APR-1999; 990S-0130510.

PR 30-APR-1999; 990S-0131449.

PR 04-MAY-1999; 990S-0132048.

PR 05-MAY-1999; 990S-0132407.

PR 06-MAY-1999; 990S-0132485.

PR 07-MAY-1999; 990S-0132487.

PR 11-MAY-1999; 990S-0132863.

PR 14-MAY-1999; 990S-0134256.

PR 14-MAY-1999; 990S-0134218.

PR 14-MAY-1999; 990S-0134221.

PR 19-MAY-1999; 990S-0134370.

PR 20-MAY-1999; 990S-0134768.

PR 21-MAY-1999; 990S-0134941.

PR 24-MAY-1999; 990S-0135124.

PR 25-MAY-1999; 990S-0135353.

PR 27-MAY-1999; 990S-0136021.

PR 28-MAY-1999; 990S-0136782.

PR 01-JUN-1999; 990S-0137222.

PR 03-JUN-1999; 990S-0137528.

PR 04-JUN-1999; 990S-0137502.

PR 07-JUN-1999; 990S-0137724.

PR 08-JUN-1999; 990S-0138094.

PR 10-JUN-1999; 990S-0138540.

PR 10-JUN-1999; 990S-0138847.

PR 14-JUN-1999; 990S-0139110.

PR 16-JUN-1999; 990S-0139452.

PR 17-JUN-1999; 990S-0139453.

PR 18-JUN-1999; 990S-0139492.

PR 18-JUN-1999; 990S-0139454.

PR 18-JUN-1999; 990S-0139455.

PR 18-JUN-1999; 990S-0139456.

PR 18-JUN-1999; 990S-0139457.

PR 18-JUN-1999; 990S-0139458.

PR 18-JUN-1999; 990S-0139459.

PR 18-JUN-1999; 990S-0139460.

PR 18-JUN-1999; 990S-0139461.

PR 18-JUN-1999; 990S-0139462.

PR 18-JUN-1999; 990S-0139463.

PR 18-JUN-1999; 990S-0139750.

PR 18-JUN-1999; 990S-0139763.

PR 21-JUN-1999; 990S-0139817.

PR 22-JUN-1999; 990S-0139899.

PR 23-JUN-1999; 990S-0140353.

PR 24-JUN-1999; 990S-0140354.

PR 28-JUN-1999; 990S-0140695.

PR 29-JUN-1999; 990S-0140823.

PR 30-JUN-1999; 990S-0140991.

PR 01-JUL-1999; 990S-0141287.

PR 01-JUL-1999; 990S-0141842.

PR 02-JUL-1999; 990S-0142154.

PR 06-JUL-1999; 990S-0142055.

PR 08-JUL-1999; 990S-0142390.

PR 09-JUL-1999; 990S-0142803.

PR 12-JUL-1999; 990S-0142920.

PR 13-JUL-1999; 990S-0142977.

PR 14-JUL-1999; 990S-0143542.

PR 15-JUL-1999; 990S-0143624.

PR 16-JUL-1999; 990S-0144005.

PR 16-JUL-1999; 990S-0144086.

PR 19-JUL-1999; 990S-0144325.

PR 19-JUL-1999; 990S-0144331.

PR 19-JUL-1999; 990S-0144332.

PR 19-JUL-1999; 990S-0144334.

PR 19-JUL-1999; 990S-0144335.

PR 20-JUL-1999; 990S-0144352.

PR 20-JUL-1999; 990S-0144632.

PR 21-JUL-1999; 990S-0144814.

PR 21-JUL-1999; 990S-0145086.

PR 22-JUL-1999; 990S-0145088.

PR 22-JUL-1999; 990S-0145089.

PR 22-JUL-1999; 990S-0145087.

PR 22-JUL-1999; 990S-0145089.

PR 23-JUL-1999; 990S-0145142.

PR 23-JUL-1999; 990S-0145145.

PR 23-JUL-1999; 990S-0145218.

PR 26-JUL-1999; 990S-0145224.

PR 27-JUL-1999; 990S-0145276.

PR 27-JUL-1999; 990S-0145913.

PR 28-JUL-1999; 990S-0145918.

PR 28-JUL-1999; 990S-0145918.

PR 02-AUG-1999; 990S-0145919.

PR 02-AUG-1999; 990S-0146386.

PR 02-AUG-1999; 990S-0146388.

PR 03-AUG-1999; 990S-0146389.

PR 04-AUG-1999; 990S-0147038.

PR 04-AUG-1999; 990S-0147204.

PR 05-AUG-1999; 990S-0147302.

PR 05-AUG-1999; 990S-0147192.

PR 06-AUG-1999; 990S-0147260.

PR 06-AUG-1999; 990S-0147303.

PR 09-AUG-1999; 990S-0147416.

PR 09-AUG-1999; 990S-0147493.

PR 09-AUG-1999; 990S-0147935.

PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151055.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3 9%; Score 19; DB 21; Length 1302;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 415 CAGAAGCAGAGATTACG 433
|||||

Db 123 CAGAAGCAGAGATTACG 105
RESULT 8
AAS26580
ID AAS26580 standard; cDNA; 401 BP.
XX
AC AAS26580;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seq ID 759.
XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;
KW cytoskeletal; cardiant; vasotropic; cerebroprotective; neurotropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnerrary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin aging; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN M0200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01341.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225216.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.


```

AC AAF16080;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:515.
XX
KM Human: prostate cancer; prostate cancer antigen; detection; diagnosis;
KM neuroprotective; cytoskeletal; cardioprotective; immunomodulatory; muscular;
KM vulnary; gastrointestinal; nephrotropic; antineoplastic; gynaecological;
KM antibacterial; gene therapy; neural; immune; reproductive; renal;
KM gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KM wound; infectious disease; ss.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PK 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM:
XX
DR WPI: 2000-587513/55.
DR P-PSDB: AAB56877.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer.
XX
PS Claim 1; Page 979; 2338bp; English.
XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytoskeletal,
CC cardioprotective, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotropic, antineoplastic, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 932 BP; 248 A; 223 C; 211 G; 242 T; 8 other:
XX
Query Match 3.7%; Score 18; DB 21; Length 932;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 237 AATAAGCTAACCAGAG 254
DB 692 AATAAGCTAACCAGAG 709

```

```

KM aspc gene; feedback inhibition; amino acid synthesis;
KM composite plasmid; ss.
XX
OS Escherichia coli.
XX
PN WO8700202-A.
XX
PD 15-JAN-1987.
XX
PF 24-JUN-1986; 86WO-US01353.
XX
PR 24-JUN-1985; 85US-0747732.
XX
PA (NUTR-) NUTRASWEET CO.
XX
PI Edwards MR, Taylor PP, Hunter MG, Fotheringham IG;
XX
DR WPI: 1987-021998/03.
DR P-PSDB: AAP70751.
XX
PT Composite plasmids contg. multiple genes in transcriptional units
PT - useful for prodn. of aminoacid(s), esp. L-phenylalanine and
PT L-tyrosine
XX
PS Disclosure; Page 25; 57pp; English.
XX
CC This sequence may be inserted into a composite plasmid and used
CC for the production of amino acids. See also AAN71053-55,
CC AAN71107, AAN71109-11 and AAP70696-97 and AAP70750, AAP70752-54.
XX
SQ Sequence 1293 BP; 322 A; 331 C; 343 G; 297 T; 0 other:
XX
Query Match 3.7%; Score 18; DB 8; Length 1293;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 132 GCTGTTGGTAAAGGTAG 149
DB 254 GCTGTTGGTAAAGGTAG 271

```

RESULT 11
AAV40259
ID AAV40259 standard; DNA; 1331 BP.
XX
AC AAV40259;
XX
DT 13-OCT-1998 (first entry)
XX
DE Escherichia coli aspc gene.
XX
KM Brevibacterium lactofermentum; lysc; L-lysine; coryneform bacterium;
KM aspartokinase; feedback inhibition; dihydrodipicolinate reductase;
KM diaminopimelate decarboxylase; aspartate aminotransferase; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 10..1197
FT /tag= "a
FT /product= "aspc"
XX
PN EP854189-A2.
XX
PD 22-JUL-1998.
XX
PF 05-DEC-1997; 97EP-0121443.
XX
PR 05-DEC-1996; 96JP-0325659.
XX
PA (AJIN) AJINOMOTO CO INC.
XX
DE Araki M, Nakamatsu T, Sugimoto M, Yoshihara Y;

XX WPI: 1998-379060/33.
 DR P-PSDB: AAM69553.
 XX
 PT Recombinant DNA autonomously replicable in coryneform bacteria -
 PT di:hydroxycinnate reductase and synthase and di:amino-pimelate
 PT decarboxylase
 XX
 PS Example 5: Page 37-38; 59pp; English.
 CC The present invention describes a recombinant DNA autonomously
 CC replicable in cells of coryneform bacteria (CB) comprising a DNA
 CC by L-lysine and L-threonine is desensitized a DNA feedback inhibition
 CC dihydrodipicolinate synthase (DHPS), a DNA sequence coding for
 CC diaminopimelate decarboxylase (DAMD) and a DNA sequence coding for
 CC aspartate aminotransferase (AAT). The present sequence encodes
 CC by CB. The L-lysine produced can be used for improving L-lysine productivity
 CC
 SQ Sequence 1331 BP; 330 A; 340 C; 350 G; 311 T; 0 other;
 Query Match 3.7%; Score 18; DB 19; Length 1331;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 132 GCTGTTGGTAAAGCTAG 149
 DB 252 GCTGTTGGTAAAGCTAG 269

RESULT 12
 ID AAN71109 standard; DNA; 3659 BP.
 AC AAN71109;
 XX
 DT 01-JAN-1980 (first entry)
 DE
 XX pheA aroF aspc operon in plasmid pME219.
 DE
 XX alpha-amyrase; feedback inhibition; amino acid synthesis;
 KW composite plasmid; ss.
 OS Bacillus licheniformis.
 XX
 FH Key
 FT CDS Location/Qualifiers
 FT 69..1241
 FT CDS /*tag= a
 FT 1259..2329 /*tag= b
 FT CDS 2344..3534 /*tag= c

W08700202-A.
 PD 15-JAN-1987.
 XX
 PF 24-JUN-1986; 86WO-US01353.
 XX
 PR 24-JUN-1985; 85US-0747732.
 XX
 PA (NUTR-) NUTRASWEET CO.
 PI Edwards MR, Taylor PP, Hunter MG, Rotheringham IG;
 DR WPI: 1987-021998/03.
 DR P-PSDB: AAP70752, AAP71677-8.
 XX
 PT Composite plasmids contg. multiple genes in transcriptional units

PT - useful for prodn. of aminoacid(s), esp. L-phenylalanine and
 PT L-tyrosine
 XX
 PS Disclosure: Page 38; 57pp; English.
 CC
 CC This sequence may be inserted into a composite plasmid and used
 CC for the production of amino acids. See also AAN71053-55,
 CC AAN71107, AAN71109, AAN71111 and AAP70656-97 and AAP70750, AAP70752 and
 CC AAP70754.
 SQ Sequence 3659 BP; 936 A; 911 C; 947 G; 865 T; 0 other;
 Query Match 3.7%; Score 18; DB 8; Length 3659;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 132 GCTGTTGGTAAAGCTAG 149
 DB 2586 GCTGTTGGTAAAGCTAG 2603

RESULT 13
 ID AAQ90655 standard; cDNA; 3776 BP.
 AC AAQ90655;
 XX
 DT 11-NOV-1995 (first entry)
 DE
 XX Eph-related PTK Cdk9 cDNA.
 DE
 XX Cdk9; Eph; protein tyrosine-kinase; PTK; cancer; diagnosis;
 KW prognosis; ss.
 OS Gallus sp.
 XX
 FH Key
 FT CDS Location/Qualifiers
 FT 290..3311
 FT CDS /*tag= a

W09515375-A.
 PD 08-JUN-1995.
 XX
 PF 07-SEP-1994; 94WO-US10140.
 XX
 PR 03-DEC-1993; 93US-0162809.
 XX
 PA (LJOL-) LA JOLLA CANCER RES FOUND.
 XX
 PI Pasquale EB, Sajjadi FG;
 DR WPI: 1995-215256/26.
 DR P-PSDB: AAR75707.
 XX
 PT Eph-related protein tyrosine kinase(s) - for monitoring and diagnosing
 PT cancer.
 PS Claim 2: Page 50-53; 129pp; English.
 CC Novel Eph-related PTK cDNA clone Cdk9 (given in AAQ90655) was
 CC isolated from a chick embryo library in lambda. Cdk9 protein
 CC (AAR75707) is closely related to Cdk5 (AAR75712). In adult tissues,
 CC Cdk9 expression is predominant in the thymus and detectable in
 CC brain, retina, kidney, lung and heart.
 SQ Sequence 3776 BP; 1005 A; 881 C; 941 G; 949 T; 0 other;
 Query Match 3.7%; Score 18; DB 16; Length 3776;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 165 GGCATTGGTGGTCTGT 182

Db 271 GGCCATTGCTGCTGT 288

RESULT 14

AA13059 standard; DNA; 9813 BP.

AA13059;

19-MAR-1999 (first entry)

Enterococcus faecalis genome contig SEQ ID NO:122.

Enterococcus faecalis; contig; detection; Enterococcal infection;

vacine; attenuation; computer readable medium; ds.

Enterococcus faecalis.

WO985055-A2.

12-NOV-1998.

04-MAY-1998; 98MO-US08985.

14-NOV-1997; 97US-0066009.

06-MAY-1997; 97US-0044031.

16-MAY-1997; 97US-0046655.

(HUMA-) HUMAN GENOME SCI INC.

Barash SC, Dillon PJ, Kunsch CA;

WPI; 1999-045171/04.

New isolated Enterococcus faecalis polynucleotides and polypeptides

used to develop products for the detection of Enterococcus and for

use in vaccines for prevention or attenuation of Enterococcus

infection.

Claim 1; Page 734-739; 2084pp; English.

A computer readable medium has been developed which has recorded on it

982 nucleotide sequences isolated from the Enterococcus faecalis genome.

AA13238 to AA13919 represent these nucleotide sequences which are

primary nucleotide sequences, also known as contigs. The computer-based

system can identify fragments of the Enterococcus faecalis genome with

commercial importance. The products can be used to detect the presence

of Enterococcus faecalis in samples. They can also be used for

diagnosing Enterococcal infection in an animal and monitoring

progression of disease, and for identifying agents which can be used to

modulate the growth or pathogenicity of Enterococcus faecalis, or

another related organism, in vivo or in vitro. In particular the

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences

can be used in vaccines to prevent or attenuate an Enterococcal

infection.

Sequence 9813 BP; 3233 A; 1698 C; 2161 G; 2719 T; 2 other;

Query Match 3.7%; Score 18; DB 20; Length 9813;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2473 ATTATATCACTGCTTT 2490

RESULT 15

ABL10082 standard; CDNA; 10432 BP.

ABL10082;

26-MAR-2002 (first entry)

Drosophila melanogaster expressed polynucleotide SEQ ID NO 24728.

Drosophila; developmental biology; cell signalling; insecticide;

pharmaceutical; gene; ss.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.

11-JUL-2000; 2000US-0614150.

(PEKE) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

WPI; 2001-656860/75.

P-PSDB; ABB65979.

New isolated nucleic acid detection reagent for detecting 1000 or more

genes from Drosophila and for elucidating cell signalling and cell-cell

interactions -

Claim 1; SEQ ID NO 24728; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent

capable of detecting 1000 or more genes from Drosophila. The invention is

useful in developmental biology and in elucidating cell signalling and

cell-cell interactions in higher eukaryotes for the development of

insecticides, therapeutics and pharmaceutical drugs. The invention

discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

sequences (ABL01840-ABL16175) and the encoded proteins

(ABB57737-ABB72072).

The sequence data for this patent did not form part of the printed

specification, but was obtained in electronic format directly from WIPO

at ftp.wipo.int/pub/published_pcl_sequences.

Sequence 10432 BP; 2894 A; 2225 C; 2170 G; 3143 T; 0 other;

Query Match 3.7%; Score 18; DB 23; Length 10432;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 170 TTGGTGTCTGCTTTTGG 187

7018 TTGGTGTCTGCTTTTGG 7035

Search completed: October 27, 2002, 20:42:22

Job time : 142.109 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 18:35:09 ; Search time 940.383 Seconds
(without alignments)
10815.053 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486
Sequence: 1 atgaacagagatgttgcga.....aagtcattccaacagaaaa 486

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
------------	-------------	--------	-------	-------------

1	486	100.0	489	6	AX252413	AX252413 Sequence
2	486	100.0	4983	1	AF184152	AF184152 Piscirickettsia
3	21	4.3	121922	1	AP000576	AP000576 Homo sapi
4	21	4.3	146327	9	AL162253	AL162253 Human DNA
5	21	4.3	160394	9	AC012273	AC012273 Homo sapi
6	21	4.3	161582	9	AP001636	AP001636 Homo sapi
7	21	4.3	161586	2	AC090309	AC090309 Homo sapi
8	21	4.3	165208	2	AC098770	AC098770 Homo sapi
9	21	4.3	183317	2	CNS07EPI	CNS07EPI Oryza sat
10	21	4.3	191454	2	AC093459	AC093459 Homo sapi
11	21	4.3	192810	2	AL626767	AL626767 Mus muscu
12	21	4.3	252571	2	AC098728	AC098728 Mus muscu
13	12	3.875	3875	3	DP022357	DP022357 Drosophila
14	20	4.1	26058	2	AC110102	AC110102 Rattus no
15	20	4.1	28798	1	U95370	U95370 Bacillus th
16	20	4.1	31147	2	AC018318	AC018318 Drosophila
17	20	4.1	32386	1	BLAU5061	BLAU5061 Bacillus
18	20	4.1	51409	1	AF024666	AF024666 Corynebact
19	20	4.1	72732	3	AC004715	AC004715 Drosophila
20	20	4.1	151673	10	AC087102	AC087102 Rattus no
21	20	4.1	160970	9	AC010138	AC010138 Homo sapi
22	20	4.1	180633	3	AC093549	AC093549 Drosophila
23	20	4.1	181654	9	AF411058	AF411058 Homo sapi
24	20	4.1	196247	2	AC104137	AC104137 Homo sapi
25	20	4.1	274307	3	AE003561	AE003561 Drosophila
26	19	3.9	4194	8	AJ1KL5079	AJ1KL5079 Kluyverom
27	19	3.9	5480	10	RATRP15	RATRP15 Rat salivary
28	19	3.9	37991	9	AP000565	AP000565 Homo sapi
29	19	3.9	39567	2	AC107960	AC107960 Homo sapi
30	19	3.9	56538	2	AC092561	AC092561 Clona sav
31	19	3.9	92693	9	AL157766	AL157766 Human DNA
32	19	3.9	100109	2	AC095401	AC095401 Rattus no
33	19	3.9	117807	2	AC106566	AC106566 Rattus no
34	19	3.9	119485	9	AC084274	AC084274 Mus muscu
35	19	3.9	121500	9	AC007916	AC007916 Homo sapi
36	19	3.9	130621	2	AC107870	AC107870 Homo sapi
37	19	3.9	131892	2	AB038490	AB038490 Homo sapi
38	19	3.9	150001	9	AC006063	AC006063 Homo sapi
39	19	3.9	160169	2	AC016660	AC016660 Homo sapi
40	19	3.9	160666	2	AC105755	AC105755 Homo sapi
41	19	3.9	167005	9	AL158070	AL158070 Human DNA
42	19	3.9	169856	9	AL591944	AL591944 Human DNA
43	19	3.9	173646	2	AC025853	AC025853 Homo sapi
44	19	3.9	174019	2	AC025734	AC025734 Homo sapi
45	19	3.9	174130	2	AC108026	AC108026 Homo sapi

ALIGNMENTS

RESULT 1
AX252413
LOCUS AX252413 489 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 5 from Patent WO0168865.
AX252413
ACCESSION AX252413
VERSION GI:15985721
KEYWORDS
SOURCE
ORGANISM
Piscirickettsia salmonis.
Piscirickettsia salmonis
Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;
Piscirickettsia.

REFERENCE 1 (bases 1 to 489)

AUTHORS Simard,N., Brouwers,H., Jones,S.F., Griffiths,S., Valenzuela,P. and Burzio,L.

TITLE Fish vaccine against piscirickettsia salmonis
JOURNAL Patent: WO 0168865-A 5 20-SEP-2001;
Aqua Health (Europe) Limited (GB)

FEATURES
source
1..489
location/Qualifiers

BASE COUNT 139 a 79 c 144 g 127 t
ORIGIN
/organism="Piscirickettsia salmonis"
/db_xref="taxon:1238"

[illegible]

```
Query Match      4.3%; Score 21; DB 2; Length 121922;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      34 ATTATTATCACTGCGTGTTTA 54
          ||| ||||| ||||| ||||| |||||
Db       70697 ATTATTATCACTGCGTGTTTA 70717
```

RESULT 4	AL162253	LOCUS	AL162253	Human DNA sequence from clone Rpl1-574F11 on chromosome 9, complete
DEFINITION	AL162253		146327 bp	-DNA
ACCESSION	AL162253			linear
VERSION	AL162253.17			PRI 21-MAR-2001
KEYWORDS	HTG		GI:13677203	
SOURCE	human			

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 146327)
Clark, G.
Direct Submission
Submitted (no)

Amniota; Craniata; Vertebrata; Euteleostomi;
Eutheria; Primates; Catarrhini; Homidae; Homo.

COMMENT

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate

30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep. This sequence was generated from part of bacterial clone c01p15 of human chromosome 9, constructed by the Sanger Centre chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>. RPl1-574F11 is from the library RPl1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>.
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RPl1-574F11. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RPl1-574F11 is at 146327 in this sequence. The true left end of clone RPl1-635M21 is at 62348 in this sequence. The true right end of clone RPl1-12D24 is at 100 in

```
misc-feature      clone_id="RPCR-11.2"
                  84304..84305
                  /note="Sequence from overlapping clone ba635n21
                  data."
misc-feature      84304..84306
                  /note="Simple clone region. Assembly confirmed by
                  restriction digest"
BASE COUNT      44360 a 29767 c 23483 g 42717 t
ORIGIN
```

	Query Match	Score 21:	DB 9:	Length 146327:
Best Local Similarity	100.0%	Pred. No. 4 9;		
Matches 21: Conservative	0;	Mismatches	0;	Indels 0;
Gaps	0;			
Db	96974	AGGCGAAGCAACAGATTTA	96994	

RESULT 5	AC012273	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	AC012273		Human sapiens clone RP11-1K7, WORKING DRAFT SEQUENCE.	AC012273	3	GI:14595892	HTGS_PHASE1; HTGS_DRAFT; HTGS_PULLTOP.	Human sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 160394)
Homo sapiens, clone RP11-1K7
Unpublished
2 (bases 1 to 160394)
Birren, R. 1990

Craniata; Vertebrata; Euteleostomi;
Catarrhini; Hominiidae; Homo.

Birren, R., Linton, L., Nusbaum, C. and Lander, E.

AUTHORS

² (cases 1 to 160394)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Brown, A., Baldwin, J., Barina, N., Beckerly, R., Boguslavsky, L., Borkgatter, B., Cooke, P., Castle, A., Colangelo, K., Collins, S., Collings, L., Collingwater, B., Ferrelle, P., DeBellano, K., Dewar, K., Domino, M., Domlan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Hearford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lebecky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (22-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 4, 2001 this sequence version replaced gi:13940700.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu
Project Information

Center project name: L1474
Center clone name: L_K_7

Summary Statistics

Sequencing vector: pM3: M77815: 24% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 159424 bases at least Q40
Consensus quality: 159851 bases at least Q30
Consensus quality: 160001 bases at least Q20
Insert size: 162000; agarose-ftp
Insert size: 160094; sum-of-contigs
Quality coverage: 14.6 in Q20 bases; agarose-ftp
Quality coverage: 14.6 in Q20 ba.
NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 16762: contig of 16762 bp in length
* 16763 16862: gap of 100 bp
* 16863 20224: contig of 3362 bp in length
* 20225 20324: gap of 100 bp
* 20325 97675: contig of 77351 bp in length
* 97676 97775: gap of 100 bp
* 97776 160394: contig of 62619 bp in length.

FEATURES

source

1. 160394
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-1K7"
/clone_1fb="RP11 Human Male BAC"
1. 16762
/note="assembly_fragment"

misc_feature
1. 16762
/note="assembly_fragment"

misc_feature
16863..20224
/note="assembly_fragment"

misc_feature
20325..97675
/note="assembly_fragment"

misc_feature
97776..160394
/note="assembly_fragment"

vector_end:T7
clone_end:T7
vector_side:right"

BASE COUNT 49405 a 32013 c 31371 g 47305 t 300 others

ORIGIN

Query Match 4.3%; Score 21; DB 2; Length 160394;

Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 240 AAAGCTAAACCGAGCTTTGGA 260
Db 81250 AAAGCTAAACCGAGCTTTGGA 81270

RESULT 6

LOCUS

AP001636/c 161582 bp DNA linear PRI 08-DEC-2001

DEFINITION
Homo sapiens genomic DNA, chromosome 11q, clone:RP11-142C4, complete sequence.

ACCESSION

AP001636 GI:17425232

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens DNA, clone:RP11-142C4.
Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (05-APR-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@psc.riken.go.jp, URL:http://npg.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
On Dec 7, 2001 this sequence version replaced gi:9927280.
Location/Qualifiers
1. 161582
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-142C4"

COMMENT

FEATURES

SOURCE

BASE COUNT 48492 a 31636 c 32205 g 49249 t

ORIGIN

Query Match 4.3%; Score 21; DB 9; Length 161582;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 ATTATTATCACTGCTTTTGA 54
Db 81922 ATTATTATCACTGCTTTTGA 81902

RESULT 7

LOCUS

AC090309/c 161586 bp DNA linear HTG 01-AUG-2001

DEFINITION

Homo sapiens chromosome 11 clone RP11-142C4 map 11, *** SEQUENCING IN PROGRESS ***; 1 ordered pieces.

ACCESSION

AC090309 GI:15055316

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 161586)
Unpublished
Homo Sapiens chromosome 11, clone RP11-142C4
Bairn, N., Bastien, V., Boguslavsky, L., Bouhgalier, B., Brown, A., Camarata, J., Campopiano, A., Chospel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, D.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glnde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,

Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., MacDonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKenna, K., McPheters, R., Melgrim, J., Menus, L., Milova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., North, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phukhng, P., Pierre, N., Pollara, V., Raymond, C., Retta, M., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Severy, P., Tlivers, M., Travis, N., Triggillo, J., Vassiliev, H., Videl, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 1, 2001 this sequence version replaced g1:14595864.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MITR

Web site: http://www.seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center Clone name: 142_C_4

* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved
1 161586 contig of 161586 bp in length.
Location/Qualifiers
1. 161586
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11"
/clone="RP11-142C4"
/clone_id="RP11 Human Male BAC"

BASE COUNT 48496 a 31637 c 32204 g 49249 t

ORIGIN

Query Match 4.3% Score 21; DB 2; Length 161586;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 81924 ATATATACAGTCGTGTTTAA 54
ATATATACAGTCGTGTTTAA 81904

RESULT 8
AC098770/c 165208 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-113M1, *** SEQUENCING IN PROGRESS
AC098770
VERSION AC098770.2 GI:117974352
KEYWORDS HTG; HTGS PHASE1.
SOURCE Norway rat
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 165208)

Austyn, D.M., Adams, C., Adlo-Oduola, B., Ali-Usman, F.R., Allen, C., Alshrook, S.L., Amaralunge, H.C., Aze, J.R., Banks, T., Babatia, J., Benton, J., Blinage, K., Blankenburg, K., Bonin, D., Bouck, J., Bowe, S., Briveta, M., Brown, E., Brown, M., Bryant, N.P., Bui, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.P., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Day, Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frenzt, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hayes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hughes, M., Holloway, C., Hollins, B., Howsi, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolyet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, B., King, L., Korvan, J., Kovar, C., Kratovic, J., Kuresh, S., Landry, N., Leal, B., Lewis, L., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, M., Massey, M., Martin, R., Martindale, A., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, M., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkwo, S., Ogburn, A., Okunoda, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, J., Peters, L., Pickens, R., Prins, E., Pu, L.L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Robjokken, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Stinson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Tang, H., Sutton, A., Swatek, A., Tabor, P., Tamerisa, K., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalona, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, A., Weinstein, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

COMMENT

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc@bcm.tmc.edu
Project Information
Center project name: G17J
Center clone name: CH230-113M1
Summary Statistics
finphrapblast
Consensus quality: 136180 bases at least Q40
Consensus quality: 144699 bases at least Q30
Estimated insert size: 151851 bases at least Q30
Quality coverage: 138535; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agorose-ip estimation
Quality coverage: 2.4x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 3812: contig of 3812 bp in length
* 3813
* 3912: gap of unknown length
* 10888: contig of 6976 bp in length
* 10889
* 10988: gap of unknown length
* 10989
* 16417: contig of 5429 bp in length
* 16418
* 16517: gap of unknown length
* 16518
* 21069: contig of 4552 bp in length
* 21070
* 21169: gap of unknown length
* 21170
* 26142: contig of 4973 bp in length
* 26143
* 26242: gap of unknown length
* 31992: contig of 5750 bp in length
* 31993
* 32092: gap of unknown length
* 32093
* 37667: contig of 5575 bp in length
* 37668
* 37667: gap of unknown length
* 37768
* 43027: contig of 5260 bp in length
* 43028
* 43127: gap of unknown length
* 43128
* 45415: contig of 2288 bp in length
* 45416
* 45515: gap of unknown length
* 45516
* 50730: contig of 5215 bp in length
* 50830: gap of unknown length
* 50831
* 55339: contig of 4409 bp in length
* 55340
* 55339: gap of unknown length
* 55340
* 59067: contig of 3728 bp in length
* 59068
* 59167: gap of unknown length
* 59168
* 63589: contig of 4422 bp in length
* 63590
* 63689: gap of unknown length
* 63690
* 66841: contig of 3152 bp in length
* 66842
* 66941: gap of unknown length
* 66942
* 69712: contig of 2771 bp in length
* 69713
* 69812: gap of unknown length
* 69813
* 74042: contig of 4230 bp in length
* 74043
* 74142: gap of unknown length
* 77230: contig of 3088 bp in length
* 77231
* 77330: gap of unknown length
* 77331
* 79217: contig of 1887 bp in length
* 79218
* 79317: gap of unknown length
* 79318
* 81009: contig of 1692 bp in length
* 81010
* 81109: gap of unknown length
* 83324: contig of 2215 bp in length
* 83325
* 83424: gap of unknown length
* 83425
* 87388: contig of 3964 bp in length
* 87389
* 87488: gap of unknown length
* 87489
* 91076: contig of 3588 bp in length
* 91077
* 91176: gap of unknown length
* 94269: contig of 3093 bp in length
* 94270
* 94369: gap of unknown length
* 94370
* 97393: contig of 3024 bp in length
* 97394
* 97493: gap of unknown length
* 97494
* 100987: contig of 3494 bp in length
* 100988
* 101087: gap of unknown length
* 101088
* 103473: contig of 2386 bp in length
* 103474
* 103573: gap of unknown length
* 103574
* 106431: contig of 2858 bp in length
* 106432
* 106531: gap of unknown length
* 106532
* 110142: contig of 3611 bp in length
* 110143
* 110242: gap of unknown length
* 110244
* 112160: contig of 1918 bp in length
* 112161
* 112260: gap of unknown length
* 112261
* 113968: contig of 1708 bp in length
* 113969
* 114068: gap of unknown length
* 114069
* 11615: contig of 2547 bp in length
* 11616
* 116715: gap of unknown length
* 116716
* 118641: contig of 1926 bp in length
* 118642
* 118741: gap of unknown length
* 118742
* 120498: contig of 1757 bp in length
* 120499
* 120598: gap of unknown length
* 120599
* 122341: contig of 1743 bp in length
* 122342
* 122441: gap of unknown length
* 122442
* 124627: contig of 2186 bp in length

* 124628 124727: gap of unknown length
* 124728 126815: contig of 2088 bp in length
* 126816 126915: gap of unknown length
* 126916 128105: contig of 1190 bp in length
* 128106 128205: gap of unknown length
* 128206 130422: contig of 2217 bp in length
* 130423 130522: gap of unknown length
* 130523 131813: contig of 1291 bp in length
* 131814 131913: gap of unknown length
* 131914 132995: contig of 1082 bp in length
* 132996 133095: gap of unknown length
* 133096 134282: contig of 1187 bp in length
* 134283 134382: gap of unknown length
* 134383 135974: contig of 1592 bp in length
* 135975 136074: gap of unknown length
* 136075 137661: contig of 1587 bp in length
* 137662 137761: gap of unknown length
* 137762 139435: contig of 1674 bp in length
* 139436 139535: gap of unknown length
* 139536 141435: contig of 1900 bp in length
* 141436 141535: gap of unknown length
* 141536 142773: contig of 1238 bp in length
* 142774 142873: gap of unknown length
* 142874 144582: contig of 1709 bp in length
* 144583 144682: gap of unknown length
* 144683 145834: contig of 1152 bp in length
* 145835 145934: gap of unknown length
* 145935 147137: contig of 1203 bp in length
* 147138 147237: gap of unknown length
* 147238 148618: contig of 1381 bp in length
* 148619 148718: gap of unknown length
* 148719 150391: contig of 1673 bp in length
* 150392 150491: gap of unknown length
* 150492 151647: contig of 1156 bp in length
* 151648 151747: gap of unknown length
* 151748 153059: contig of 1312 bp in length
* 153060 153159: gap of unknown length
* 153160 154185: contig of 1026 bp in length
* 154186 154285: gap of unknown length
* 154286 155334: contig of 1249 bp in length
* 155335 155634: gap of unknown length
* 155635 157043: contig of 1409 bp in length
* 157044 157143: gap of unknown length

Query Match 4.3%: Score 21: DB 2: Length 165208:
Best Local Similarity 100.0%: Pred. No. 4.9:
Matches 21: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
Oy 235 AAAATTAAGCTTAACCCAGAGT 255
Db 37871 AAAATTAAGCTTAACCCAGAGT 37851
|||||
CNS07EFL
LOCUS
DEFINITION
ORYZA SATIVA CHROMOSOME 12 CLONE OSNBA0009F13, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION
AL513003
VERSION
AL513003.1 GI:12329144
KEYWORDS
HTG: HTGS, PHASE2.
SOURCE
ORYZA SATIVA.
ORGANISM
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 183317)
Salse,J., Choisne,N., Orjeda,G., Regad,F., Lortieux,M., Cooke,R.,
Deiseny,M., Robert,C., Broctier,P., Wincker,P., Cruaud,C.,
Artiguenave,F., Saurin,W., Salanoubat,M., Quetier,F. and
Weissenbach,J.
Oryza sativa chromosome 12 sequencing
JOURNAL
Unpublished

REFERENCE 2 (bases 1 to 183317)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2001) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT

- Web : www.genoscope.cns.fr
- IMPORANT : This sequence is unfinished and does not necessarily represent the correct sequence
Work on the sequence is in progress and the release of this data is from E.coli, yeast, vector, phage, etc.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
the accession number will be preserved.

FEATURES

Location/Qualifiers

1..183317
/organism="Oryza sativa"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:4530"
/chromosome="12"
/clone="OSJNBa009p13"
/clone_idb="CUG1 Nipponbare BAC"

BASE COUNT

52033 a 38000 c 38408 g 54873 t 3 others

Query Match

Best Local Similarity 100.0%; Score 21; DB 2; Length 183317;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 TTGGAAAGGTAAACGAGG 276
Db 67144 TTGGAAAGGTAAACGAGG 67124

RESULT 10
AC093459/c

LOCUS

AC093459 191454 bp DNA linear HTG 07-FEB-2002

DEFINITION

Homo sapiens chromosome 2 clone RP11-499M11, WORKING DRAFT
SEQUENCE, 3 unordered pieces.

ACCESSION

AC093459.4 GI:18584937

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 191454)
Waterston, R.H.
The sequence of Homo sapiens clone

AUTHORS

Waterston, R.H.

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 191454)

AUTHORS

Waterston, R.H.

JOURNAL

Submitted (25-AUG-2001) Genoscope Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Feb 7, 2002 this sequence version replaced g1.18308909.

COMMENT

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.edu
Project Information
Center project name: H.NH0499M11
----- Summary Statistics -----
Sequencing vector: M13; 58
Sequencing method: plasmid; 948
Chemistry: Dye primer ET; 0% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180361 bases at least Q40
Consensus quality: 180903 bases at least Q40
Insert size: 184000; agarose-ff
Insert size: 191254; sum-of-contigs
Quality coverage: 8.23 in Q20 bases; agarose-ff
Quality coverage: 7.92 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 9927 10026: contig of 9926 bp in length
* 10027 75543: gap of unknown length
* 75544 75644: contig of 6517 bp in length
* 75644 101454: gap of unknown length
Location/Qualifiers
1..191454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-499M11"
1..9926
/note="assembly_name:Contig1"
10027..75543
/note="assembly_name:Contig17"
75644..191454
/note="assembly_name:Contig18"

FEATURES

Source

1..191454
Location/Qualifiers

misc_feature

1..9926
/clone="RP11-499M11"

misc_feature

10027..75543
/note="assembly_name:Contig1"

misc_feature

75644..191454
/note="assembly_name:Contig17"

BASE COUNT

57526 a 36639 c 38791 g 58298 t 200 others

Query Match

Best Local Similarity 100.0%; Score 21; DB 2; Length 191454;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

240 AAAGCTAAACGAGTTTGA 260

Db 127210 AAAGCTAAACGAGTTTGA 127190

RESULT 11
AL626767/c

LOCUS

AL626767 192810 bp DNA linear HTG 12-OCT-2001

DEFINITION

Mus musculus chromosome 4 clone RP23-421C7, *** SEQUENCING IN
PROGRESS ***

ACCESSION

AL626767.2 GI:16073773

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 192810)
Burton, J.
Direct Submission

AUTHORS

Burton, J.

JOURNAL

Submitted (10-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Oct 11, 2001 this sequence version replaced g1.16031460.

COMMENT

Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk
Project Information
Center project name: BM421C7


```

----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid: L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 187761 bases at least Q40
Consensus quality: 186634 bases at least Q40
Consensus quality: 189284 bases at least Q20
Insert size: 190910; sum-of-contigs
Insert size: 213770; 3.4% error; agarose-fp
Quality coverage: 8.22x in Q20 bases; sum-of-contigs Quality
Coverage: 7.45x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source
    location/Qualifiers
    1..192810
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /chromosome="4"
    /clone="RP23-421c7"
    /clone_1id="RPC1-23"
    1..15994
    /note="assembly_fragment:02908
    fragment_chain:1
    clone_end:SP6
    vector_side:left"
    16095..26569
    /note="assembly_fragment:00217
    fragment_chain:1"
    26670..52345
    /note="assembly_fragment:00503
    fragment_chain:1"
    52446..57070
    /note="assembly_fragment:03381
    fragment_chain:1"
    57171..60077
    /note="assembly_fragment:02852
    fragment_chain:1"
    60178..62198
    /note="assembly_fragment:02444
    fragment_chain:1"
    62269..66215
    /note="assembly_fragment:00203
    fragment_chain:2"
    66316..93884
    /note="assembly_fragment:03792
    fragment_chain:2"
    93985..103843
    /note="assembly_fragment:03646
    fragment_chain:2"
    103944..116303
    /note="assembly_fragment:05128
    fragment_chain:2"
    116404..120297
    /note="assembly_fragment:00677
    fragment_chain:3"
    120398..124934
    /note="assembly_fragment:02132
    fragment_chain:3"
    125035..128604
    /note="assembly_fragment:03420
    fragment_chain:3"
    128705..131472
    /note="assembly_fragment:03428
    fragment_chain:4"
    131573..145198
    /note="assembly_fragment:04437
    fragment_chain:4"
    145299..148627
    /note="assembly_fragment:05331"
    148728..150958

```

```

/note="assembly_fragment:00460
fragment_chain:5"
151059..161424
/note="assembly_fragment:00236
fragment_chain:5"
161525..186255
/note="assembly_fragment:04238
fragment_chain:5"
186356..192810
/note="assembly_fragment:04763
fragment_chain:5"
clone_end:T7
vector_side:right"

BASE COUNT    49437 a 45229 c 45295 g 50941 t 1908 others
ORIGIN

Query Match      4.3%; Score 21; DB 2; Length 192810;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 171 TGGTGCTCTGTTTGGCTGC 191
Db 130405 TGGTGTCTGCTGTTTGGCTGC 130385

RESULT 12
AC098728/c 252571 bp DNA linear HTG 14-NOV-2001
LOCUS      Mus musculus chromosome UNK clone RP23-3H22, WORKING DRAFT
DEFINITION
SEQUENCE 7 unordered pieces.
ACCESSION  AC098728
VERSION    AC098728.2 GI:169241170
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE    house mouse.
ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 252571)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
2 (bases 1 to 252571)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (31-OCT-2001) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Nov 14, 2001 this sequence version replaced gi:16554425.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BA0003H22
----- Summary Statistics -----
Sequencing vector: M13; %
Sequencing vector: plasmid; %
Chemistry: Dye-terminator Big Dye; % of reads
Chemistry: Dye-terminator Big Dye; % of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 249548 bases at least Q40
Consensus quality: 250157 bases at least Q30
Consensus quality: 250759 bases at least Q20
Insert size: 250000; agarose-fp
Insert size: 251971; sum-of-contigs
Quality coverage: 17.62 in Q20 bases; agarose-fp
Quality coverage: 17.54 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1055: contig of 1055 bp in length
 * 1056 1155: gap of unknown length
 * 1156 2391: contig of 1236 bp in length
 * 2392 2491: gap of unknown length
 * 2492 3645: contig of 1154 bp in length
 * 3646 3745: gap of unknown length
 * 3746 4905: contig of 1160 bp in length
 * 4906 5005: gap of unknown length
 * 5006 23951: contig of 18946 bp in length
 * 23952 130884: contig of 106833 bp in length
 * 130885 130984: gap of unknown length
 * 130985 252571: contig of 121587 bp in length.

FEATURES
 source
 1.252571
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="XNK"
 /clone="RP23-3H22"
 1. 1055
 /note="assembly_name:Contig27"
 misc_feature
 1156..2391
 /note="assembly_name:Contig28"
 misc_feature
 2492..3645
 /note="assembly_name:Contig35"
 misc_feature
 3746..4905
 /note="assembly_name:Contig37"
 misc_feature
 5006..23951
 /note="assembly_name:Contig38"
 misc_feature
 24052..130884
 /note="assembly_name:Contig39"
 misc_feature
 130985..252571
 /note="assembly_name:Contig40"

BASE COUNT 73702 a 54784 c 54357 g 69126 t 602 others
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Pred. No. 4.8;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 140043 AGTAGCTATATATACACT 140023

RESULT 13
 LOCUS DP022357 3875 bp DNA linear INV 17-APR-1997
 DEFINITION Drosophila pseudobscura runt gene, complete cds.
 ACCESSION U22357
 VERSION U22357.1 GI:722344
 KEYWORDS
 SOURCE
 ORGANISM
 Drosophila pseudobscura.
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 3875)
 Pepling,M.E. and Gerger,J.P.
 Conservation and function of the transcriptional regulatory protein
 Runt
 Proc. Natl. Acad. Sci. U.S.A. 92 (20), 9087-9091 (1995)
 9601614
 2 (bases 1 to 3875)
 Pepling,M.E.
 Direct Submission
 Submitted (07-MAR-1995) Melissa E. Pepling, Biochemistry and Cell
 Biology, SUNY at Stony Brook, 346 Life Sciences Building, Stony

Brook, NY 11794, USA
 Location/Qualifiers
 1..3875
 /organism="Drosophila pseudobscura"
 /db_xref="taxon:7237"
 gene
 prim_transcript 1368..3875
 join(1799..2549,2841..3775)
 /gene="runt"
 CDS
 join(1799..2549,2841..3775)
 /gene="runt"
 /codon_start=1
 /protein_id="AAB52378.1"
 /db_xref="GI:722345"
 /translation="MHLPNGPTWVNSTAVAHHTOVLAAAAAANSTPAPAFSA
 AOSTSLANTSTHSASSTGSSSTGSDPISATNTANTSSSSNSNTANNTANNSSNSN
 SNSNSAKMPSMTDMFASLHEMLQEVHGLAQGSIIICSLPIMHRSNKSILPGAFK
 VIALDDVDGTLVSIKCGNDENCGELRNCTITMKQVAKFNDLRFVRSRGSKFTL
 TITATYVVOIASYSKAIKTVYGPPEPRKOSYGYPHGCAFPMLNPMLDAAYMT
 YGADYPRHOAAQAAVHHHPALSKASPSSTSVSPASAAASCGANGAPADYHH
 VSGITPPSPCAPTACGPAAMPSPPGAAAPAAVAILPQPFNVHAAAAAAGAACOOOH
 AKSTPAHFHPYNFAAAGLRARNAAAVLAHGGDATGSHISPASRSSSTPTOOHV
 LKIKNTSIETSIHQASDADSDDDQIDIVASEVDLKSITSSSSSFLAHSVPLR
 MRCDLXAPSAIKPLPFHESAAAAATNRQASPEFTLLPAATKIKNSTVQOKTWRPY"

BASE COUNT 895 a 1207 c 1002 g 766 t 5 others
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2075 TGTTGTTGGCGGTTCCTG 2056

RESULT 14
 LOCUS AC110102 26058 bp DNA linear HTG 10-FEB-2002
 DEFINITION Rattus norvegicus clone CH230-209C7, *** SEQUENCING IN PROGRESS
 AC110102
 ACCESSION AC110102.1 GI:18643397
 VERSION
 KEYWORDS HTG; HTGS; PHASEL.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 26058)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Allisman,F.R., Allen,C.,
 Albrooks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbara,J., Benton,J., Bimaye,K., Blankenburg,K., Bonni,D.,
 Bouck,J., Bowler,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Bunay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Earnhart,C.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Douthwaite,K., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Fallis,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
 Gabisl,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Hayida,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Hollaway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Koryah,J., Kovar,C.,
 Kratovic,U., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,Y., Louised,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,


```

/protein_id="AAD04757.1"
/db_xref="GI:3080742"
translation="MGNTFFPLTHAORRIWYERFYPGTSVNSLSGFKKSKSGTDS
GQLEAFIRQFVRENETRPMREDEDEPKFQYIADEPFOLECPDASGSSGAGVLTW
GQAEARRPLPLPGSPLEKFAVYVRISEESSEKROKDAWNNQVETITLGRITDYLK
IQLDAARFKEIPLPGSPLEKFAVYVRISEESSEKROKDAWNNQVETITLGRITDYLK
GNTNAREKQIMGMVLTITQYCEKHNISVLSLESLITLTHYRITGOKDVLSTPM
DLDEROPHVKSLPSVLEIYQWQOKETLSLEPERFSSGVENDISIVKRMPTG
LAIDEPREBELTCEBELAVLERMTLLEDLALPQKTALEIIVAEKEVLTITG
SOTIAYDKMNTLAGLEQKADHEKTAIVYEGOKSTIRLENOENRILARRGCT
PDAPAVVNERSEKVIITAMLCVAKAGAYVDPGPEERIFMLDEKVAIVITDGS
LTPETAEVYRSEALSLENGHPSSEAGCHAYIITVSGTGRPKRMVLEHQVH
LVNGLOOAVGAVDODDLKALLAPFHDASVOGITFSLITLQGITIYVKKVMSGRAL
SOYVRRHODIVTGTDPHMLDLSADPLSGVKIRMLVGEALISVARELLDQLEAF
AESVDPVTVNGPTECVDASSFCTMNHADLOGDRAVYIPGRIQNNPTILDENAL
LPDVEGELYAGDVGRYLNPDTIADKLEDEPFVGGFMYRTGAVMLPDGTV
FTRKDQVKIRGRIELEGESVLEDEIPLTSGKVDKRALRHEVSVGAEVPLQV
EHLQGLREHLSKHLDPYNTPSHEVEDEIPLTSGKVDKRALRHEVSVGAEVPLQV
FRNEYEKMVGIMQVILGTBOVGHDOPELGSISAKAMLAKITKAGVEVPLQV
PEKPTVALSGFVSEAKDGFVIEPAPESDDYPLISAOIRITVSQLEDAVLLQV
AAALTEGTLVORLEBAFQKILDRHEALRTFTVDESEPMQNIHQVOKIEKVAOL
KPIEOMAKSFVRRTDLAKAPIRAGIVLADRRHVLTPDHMLVSDGVSITILNEAL
ALYKEBELPEPNIHTKDYAFWORTHAPKQKEKALWOSVFARELPIVOLTIDPERS
VOSEEDRYSAVLPODMKEKLAIVLQNGATVYMLAAYNMULAKYSGOEDVITGP
AAGRHSRDLSEGLMEVNTLAIRKVSGRPTPLMLAAYNMULAKYSGOEDVITGP
EKVARSRDLSEGLMEVNTLAIRKVSGRPTPLMLAAYNMULAKYSGOEDVITGP
GELLIDDLSTKFKKDIAERMLKAYIHILEGTDAPLRIGEYSILTEETPNQLEA
FNPASDVPREKTIYOLFEBOAREGSHALOEKVMVSGELNKRKNOLARLRISG
VOPGTAITLARSASENVIGILAVLAKAGAYVIDPDEKRVQHPEDGCAVLLQV
KAMKPLAEMAFGGLDLEDEHLLGDASDLRLPISPEAMANTVSGTGPCKM
VSHRNLITVAKNTYLDVMPDSILVLSINVYDAMFDEFGSILNKAQLIADKDTIL
DMSRILAHLEKEKTYLMTITVLSINVYDAMFDEFGSILNKAQLIADKDTIL
TVAGGLLHMYGSESTVTTTALPILDMRDSIKGLRRVITGEGASVDARALN
RIDQVYIRGORTELEJHNLVTHEMVOEAVALVYGRKDMCAVYTAODELISG
ELRYVAEGLRGMLPSPVMOLELPLTNGKNDKRALPEPDMOAOQEVYTAODELISG
GAOLADMOEVLNPRICVHNFELGHSILGNTILARIQOEMWMDLOLDIQAOT
IETLAQAKTEKSAVYIEAVPDROTYPVSSQKRYVLOQIEGASVYMPADLP
EGKDLAKRLESAAOMLIRHEAFRTTEITVQDGVORIVEAEELITVYIDDEKALV
LITEPIREPDLKAPLEPMSTIRFPGKELTQDGVORIVEAEELITVYIDDEKALV
EALBEPRIQYKDFSWOOHLTEHRYOEEMVLEKELAGELPVITLIPADYRPSITQTE
GSRSLFAKPELVQOOLKRLAKETDSTLYVLAASYLISGOSVYIPGSAAPR
HVDNRILIGMENVTLAIRPEBDKPFSAFLEVEKETTLAGFHEQDIPVGSAPAR
OQDMRNPLFPAVSEMRNDLKOISMEGITLPEYFAHOTKPTLITLAAEDGLLVE
LAVPODTHHEIFEAQVLRTPORGAAYVIDPSYPAERTIYLDSGAELITOTMLA
RIGIMVRELEMAAYLIRKPKGAAYVIDPSYPAERTIYLDSGAELITOTMLA
PEPSEGETLISMSLEETTXDDEVNPOAGOSDNLITVYSGTGPCKMVAHRS
LVNLCYWHNDARVTEBOKSAYAGFEGDASVEMRPLYMAGLELHIDEALRNDITR
ENNVATSTAIIDPKGMLSIGKPIANTRAVYLDONNEVOPVAGCELCIARGLARIG
LNKREPTARKEPDEPFVBERMYRTGDAVYLDONNEVOPVAGCELCIARGLARIG
FVQALRSEVOEAVIDIDPAGCNKRLCGIVVANOELTDESLARKLAOTIPRTVSE
IDNNEFALGDSISIGIOMASRLQORHMKLEMDLFOHPITIGELSAVYOQADDPIDON
MYOENGLVLIQORWFERREPTNHNHNSQVMDLALGTDEPTACKALSLIEHNDIR
EYARLEEVGSAFAKEITVQELIEQAEALPRDEADKRRHAKRAEASLEP
TEQMTVYHAYVNTBMNDLILALGALKEWNGOEDPITCLEGHREEDITIDGNISRT
VGMTSOPPALIOMHSGSDIGYOIKOIKELRHTPNKGIDYIGYRLRTEEGKAPIT
HDSFNTQFTBMASGSDIGYRTSPSGDPTETEPNMLDIVGIEGGLIYMSIAY
HSLYEKSTVAAASFKYTLQIDHCLERDGEELTPSDGLDELTLEELDKLMEITP
"
gene
11230.21996
/feature="1icB"
11230.21996
/feature="1icB"
/function="antibiotic-biosynthesis"
/note="1icB"
/codon_start=1
/transl_table=1

```

```

/product="1ichenysin synthetase B"
/protein_id="AAD04758.1"
/db_xref="GI:3080743"
translation="MGKOKTQVYVPLTMOGMLYHAMLDPESSYFTOLEFTDGEF
DLGIFEVSQVQALVRSYDILRTFVHQIQLOKROVVLAEARAYFERDLADLDEKQNN
RIDVYKQVQAAQFENLAKDMLKTAIVFLDNRKTLVWVSHHNLVMDGSMGIMKRLF
QNTAEFAFANTVPLIDGKRTADYIKMLGRQKGAASVYWHRLSYGSESVTPGKPR
PDESRNNEEFVVDKRNVDLQOANRYHVAIPNLFOIMQAVLAVYNTDYPGKPR
YELQSGRSEDGILTEHAGLFNTIPVRLKTDONASRTLRLEFKAQOHAAYARVYPL
VYGGAKMTKLLTNGAADAFAIRFOAGHGLTRIGABEALINPSKASNAAMSKOEDT
VYGMORBPAPLIMMERLSLEASMLAYKAGGVVDPDGPPEERIRPLIDPSGRT
VLTQSTOISLEGEVLAVANAAKEEDANLEHVNKDEDLAYIYTSGSTGRKVM
DOKRNEMLKLEGTITVWMLSPPLNOLAODDPGMATLNDLITGGDALVHGVNVRK
BSPELSLWNGKQEPENTTSCTCLIDQAEERTIPGRIGNSTAYIDEGDALVPKRETL
PGEUCVGDGVARGYINQPELIDKPEVPAEKRMVRTGDLAKMLDESGLOPGR
DHQVKGVPFIEGEVEALALBESIISAAVIRENNAGNETCAIYAGARPAQOLT
ELISLPEYWIPIAHLELDSMLANGVDRQULPAIERTCAIYAGARPAQOLT
IWEISLIGIKOPIDNFTSIGHSLSKAMITAKIOEOLQEVYKLEKPTIRLEAL
ERLEAERLILNRHETLSELDIGEIVOIHOVYFHTESVAEAEAAARERT
RPELNRKAVLVAELVLEERHMLIDMHHITDSSGTLIRPLAGFYGEARAP
KLHYKOPAWQNGPEHOKLEKQOBYVIDVKEKLESDGLIPDRPFRPKRSFEEDRY
FGIDKRLDIOKLESEADATLYWFLAANILLSKYASODLIVGSPVGRHYPDH
NIPGMVNTLALNRKPEKTFPOLOFQKETSIVHRRAKDLLEAVEHEDGTCSTFA
VSLFEETVFNQNMNEVPLTKLGLDLSISVHRRAKDLLEAVEHEDGTCSTFA
BRPFVDEEKADMEPPOPAVNGERTITVYRLVLEKAOILANVIRSCVQKQAELE
LNRSAEVAALILGVKSGGAFPLDPEMTEVLEKAOILANVIRSCVQKQAELE
ITGGLIDVRRDITLANSKGRPHPTIGASLAIYTSGTGTQVGLERHNRANVSV
QGINLHSSDDKTMILSYAPDLGYSFVPLVAGLELHVPETITHEATHTIGE
OQALYKMLPESLTHYVOOSPALAKGASPTIRIGPSAGALVLAHSLVAGAGE
INHYGPLETTIGAKLAPEDMDIGLOFASPTIRIGPSAGALVLAHSLVAGAGE
VLRITGGLARGYNNBELTAKFVANYTGTGTEWYPGDIVLRNNGELTEGTRDD
LVIRGRITELKTEIEAFNGLDIERAIVLSEFTDSSGLDLCAYIOAKQOLVSLRER
LSERLDVYVITBEAFNGLDIERAIVLSEFTDSSGLDLCAYIOAKQOLVSLRER
GISEEPLVQVKGDDNDFDLGHSLSKGMVRLPEQOGLNGSDSYRAPSTELKAC
PDRBRARAFADLVNRHESLRTAFLETIDGTVOIVHEKADILKVIHIREDAEELVR
EKLTPDSTAVLIRAEELISSRHLIDTHHITADVSRLSVKTIADLRKAEVLR
VRFKASDAKALIRRLAENTNTLNTVMAVNFLESLTSGODIYVGRPVOSFADR
LKDMGFAVSLAKKHVVDQSESEFELGNNLSLAADHODYPPEELIAGRTNAD
MSRDLFNWMLTTEEDPKETDLDGLTTPETISHAAKADPLIGAFEDHRIADP
YATDLFQKTIERWGSYTLNTDGLTTPETISHAAKADPLIGAFEDHRIADP
PEKTVHEIFEAQVLRTPORGAAYVIDPSYPAERTIYLDSGAELITOTMLA
IMKPSLEMAAYLIRKPKGAAYVIDPSYPAERTIYLDSGAELITOTMLA
FTGDVIDILREBSIILGAKAGAYVIDPSYPAERTIYLDSGAELITOTMLA
LVNLCYWHNDARVTEBOKSAYAGFEGDASVEMRPLYMAGLELHIDEALRNDITR
LNQYPERKNTITTELPOLCQEMLEMDLFOHPITIGELSAVYOQADDPIDON
ENNVATSTAIIDPKGMLSIGKPIANTRAVYLDONNEVOPVAGCELCIARGLARIG
FVQALRSEVOEAVIDIDPAGCNKRLCGIVVANOELTDESLARKLAOTIPRTVSE
IDNNEFALGDSISIGIOMASRLQORHMKLEMDLFOHPITIGELSAVYOQADDPIDON
MYOENGLVLIQORWFERREPTNHNHNSQVMDLALGTDEPTACKALSLIEHNDIR
EYARLEEVGSAFAKEITVQELIEQAEALPRDEADKRRHAKRAEASLEP
TEQMTVYHAYVNTBMNDLILALGALKEWNGOEDPITCLEGHREEDITIDGNISRT
VGMTSOPPALIOMHSGSDIGYOIKOIKELRHTPNKGIDYIGYRLRTEEGKAPIT
HDSFNTQFTBMASGSDIGYRTSPSGDPTETEPNMLDIVGIEGGLIYMSIAY
HSLYEKSTVAAASFKYTLQIDHCLERDGEELTPSDGLDELTLEELDKLMEITP
"
gene
22033.25901
/feature="1icC"
22033.25901
/feature="1icC"
/function="antibiotic-biosynthesis"
/note="1icC"
/codon_start=1

```

```
/tranyl_table=11
/product="lichenysin synthetase C"
/protein_id="A004759.1"
/db_xref="GI:3080744"
/tranlation="MKGDYTLISQFKKEHVODIYVLSPMQGMLEFHTLHPGQSFYIEQ
ISMOVKGSFOKDLERKSMNVYIGRYDIFRTVFEHEKMRPVQVVKERSFOAEIDLS
GLSEAEONERIEDYKRRDKKEGFNLKSDIPMTAVFKGODREYEWWSYHHILLDGCW
FGIVQELFEVYNALRENRLYSIGPYKPKDYIKWLESQDKORSILAVMEHHLAGFEGQ
TTPSEORKKPEQAGNQPHEELFTLPKEDTEVFTQLAKAHHTTLSTALQAVSVLSRY
QSRDLIFGVVSGRPADIOGVEHMGVGLFNVVPRKRVTEGAHTTFTELVAELQKQALE
AEPHQYMPLYDIOSHIAAPNLIDHIIVFENYPLQEBANKQHEEKNLGFTMDITVPEKS
NYDLNLASPGSEMLKLAVNRNVFEPSTILOLKEQLLTAIREIITNPEOPLDDICV
SKKEERLLHEFENGVOVKOTGLTIPOPEOCAEALPEOPAVSAAGRTLYRELNQEA
NKVAHLRRKNRGSEPVALLFRSPENWIAITAVLKAGGAVLPIDPEYPARIQYML
EDSGAVCMLETOELAGQAASLSFHQNTILIDPVAESGSRNLEIAEPDOLAIVMTT
```

```
Query Match 4.18; Score 20; DB 1; Length 28798;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 229 CAGCATAAATAAAGCTAAA 248
|||||
Db 27381 CAGCATAAATAAAGCTAAA 27400
```

Search completed: October 27, 2002, 21:48:35
Job time : 1558.38 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:56:01 : Search time 18.1865 Seconds
(without alignments)
1540.986 Million cell updates/sec

Title: US-09-677-374-2

Perfect score: 836
Sequence: 1 MNRGCLQGSSLIITISVFLVNG.....ITGFACRQPDGRMQUISTFK 162

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

SPTRMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	836	100.0	162	2	Q9F9K8	Q9F9K8 pisciricket
2	313.5	37.5	148	2	O54381	O54381 rickettsia
3	291	34.8	159	2	Q9F9F2	Q9F9F2 rickettsia
4	266.5	31.9	137	2	O52252	O52252 rickettsia
5	260.5	31.2	144	2	Q9K2N6	Q9K2N6 male-killin
6	259.5	31.0	137	2	Q31065	Q31065 rickettsia
7	259.5	31.0	144	2	Q9K4W8	Q9K4W8 male-killin
8	252	30.1	154	2	O53154	O53154 rickettsia
9	247	29.5	151	2	Q9F9G9	Q9F9G9 rickettsia
10	237.5	28.4	131	2	O9F001	O9F001 rickettsia
11	237.5	28.4	131	2	O52637	O52637 rickettsia
12	236.5	28.3	131	2	O9L522	O9L522 rickettsia
13	223.5	26.7	105	2	O31208	O31208 rickettsia
14	143	17.1	199	16	Q985G4	Q985G4 rhizobium
15	140	16.7	77	2	O9AGC7	O9AGC7 rickettsia
16	130.5	15.6	136	16	Q92R89	Q92R89 rhizobium

17	128.5	15.4	182	16	Q9HX13	Q9HX13 pseudomonas
18	116	13.9	154	16	O914S1	O914S1 pseudomonas
19	112	13.4	257	16	O9RGX0	O9RGX0 xyella fas
20	111	13.3	139	2	O52854	O52854 rhizobium
21	110.5	13.2	232	16	O9A3X8	O9A3X8 caulobacter
22	108.5	13.0	155	2	O9F6B1	O9F6B1 edwardsiell
23	107.5	12.9	153	2	O69776	O69776 rhizobium e
24	107.5	12.9	155	2	O9RA95	O9RA95 serratia sp
25	106.5	12.7	221	16	Q92ST9	Q92ST9 rhizobium m
26	106	12.7	155	2	O9RB08	O9RB08 pectobacter
27	105.5	12.6	223	2	Q9XC44	Q9XC44 porphyromon
28	101.5	12.1	304	16	O91762	O91762 pseudomonas
29	99.5	11.9	83	16	O92LP2	O92LP2 pseudomonas
30	99	11.8	154	16	O9CN83	O9CN83 pasteurilla
31	98	11.7	172	16	P76572	P76572 escherichia
32	98	11.7	220	2	Q9XAX8	Q9XAX8 pseudomonas
33	96	11.5	223	16	O9KSRI	O9KSRI vibrio chol
34	95.5	11.4	79	16	O92NT4	O92NT4 rhizobium m
35	94.5	11.3	838	2	Q9AL49	Q9AL49 shigella fl
36	94	11.2	608	10	Q9SUX1	Q9SUX1 arabidopsis
37	93.5	11.2	257	16	Q9A8W8	Q9A8W8 caulobacter
38	92	11.0	105	16	Q9BP93	Q9BP93 rhizobium
39	91	10.9	544	5	O46171	O46171 nephila cla
40	90	10.8	105	16	O983Y0	O983Y0 rhizobium
41	89	10.6	137	16	O9HU07	O9HU07 pseudomonas
42	88.5	10.6	161	2	O68681	O68681 bacillus me
43	88.5	10.6	216	16	Q98NA9	Q98NA9 rhizobium
44	87.5	10.5	407	16	O31557	O31557 bacillus su
45	87	10.4	201	16	O930X4	O930X4 rhizobium m

ALIGNMENTS

RESULT 1
ID Q9F9K8 PRELIMINARY: PRT: 162 AA.
AC Q9F9K8:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE 17 KDA ANTIGEN.
GN OSPA.
OS Piscirickettsia salmonis.
OC Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;
OC Piscirickettsia.
OX NCBI_TaxID=1238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LF-89;
RA Kuzuk M.A., Burton J., Thornton J.C., Kay W.W.;
RT "Identification of a genus-common Rickettsial surface antigen in the
RT salmonid pathogen Piscirickettsia salmonis.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF184152; AAC17000.1; -; DDB99E6FD94A527E CRC64;
SQ SEQUENCE 162 AA; 17661 MW; DDE99E6FD94A527E CRC64;

Query Match 100.0%; Score 836; DB 2; Length 162;
Best Local Similarity 100.0%; Pred. No. 9.6e-68;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNRGCLQGSSLIITISVFLVNGCAQNFSEVGAATGAVVGAAGLFFKSGSRVMAITGA 60
Db 1 MNRGCLQGSSLIITISVFLVNGCAQNFSEVGAATGAVVGAAGLFFKSGSRVMAITGA 60
QY 61 VLGLLSKTIQSDMDQDKIRLNOSLEKRVKAGVTRRNPDGTGSVSEVPRTQRYNKK 120
Db 61 VLGLLSKTIQSDMDQDKIRLNOSLEKRVKAGVTRRNPDGTGSVSEVPRTQRYNKK 120
QY 121 ERROQYCRFEQOKAMIGAKQOEIYGTACROPDGRMQUISTFK 162
Db 121 ERROQYCRFEQOKAMIGAKQOEIYGTACROPDGRMQUISTFK 162

[illegible]


```

SQ SEQUENCE 144 AA; 14785 MW; C8254739CA56AE7 CRC64;
Query Match 31.2%; Score 260.5; DB 2; Length 144;
Best Local Similarity 39.7%; Pred. No. 5.9e-16;
Matches 52; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

OY 26 SROEVGATGAVVGVGAGQLFGKSGRYAMAIGGAVLGILGSKIGOSMPDQDK-----IK 81
DB 17 NKQGTGTLGGAGGALLGSGFGKSGKQGLVGVGALLGAVLGGOIGAGMDQDRRLAELT 76
OY 82 LNQSLKRYKAGQVTRWRNPDTGNSYSVEPVRTYORYNKQERRQOYCREFOQKAMIAQOK 141
DB 77 SQRALEAPSGSNVWRNPDMNGHYTPNKTY-----RNSTGQYCREYTOTVVIQKQO 131
OY 142 EIVGTACRQPD 152
DB 132 KAYGNACRQPD 142

RESULT 6
O31065 PRELIMINARY; PRT; 137 AA.
AC O31065; Q9MM02;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-OCT-2001 (TREMblrel. 18, last annotation update)
DE 17 KDA ANTIGEN (17 KDA PROTEIN) (FRAGMENT).
OS Rickettsia honei.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_Taxid=37816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TT-118;
RA Billings A.N., Yu X.-J., Teel P.D., Walker D.H.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99045882; Pubmed=9828442;
RA Stenos J., Roux V., Walker D., Raoult D.;
RT "Rickettsia honei sp. nov., the aetiological agent of Filanders Island
RT spotted fever in Australia.";
RL Int. J. Syst. Bacteriol. 48:1399-1404(1998).
DR EMBL: AF027124; AAB81846.1; -.
DR EMBL: AF060706; AAD20231.1; -.
DR EMBL: AF060704; AAD20230.1; -.
FT NON_TER 1
OX NCBI_Taxid=1
SQ SEQUENCE 137 AA; 14167 MW; 75BC1D0D745B428C CRC64;

Query Match 31.0%; Score 259.5; DB 2; Length 137;
Best Local Similarity 39.7%; Pred. No. 5.9e-16;
Matches 52; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

OY 26 SROEVGATGAVVGVGAGQLFGKSGRYAMAIGGAVLGILGSKIGOSMPDQDK-----IK 81
DB 10 NKQGTGTLGGAGGALLGSGFGKSGKQGLVGVGALLGAVLGGOIGAGMDQDRRLAELT 69
OY 82 LNQSLKRYKAGQVTRWRNPDTGNSYSVEPVRTYORYNKQERRQOYCREFOQKAMIAQOK 141
DB 70 SQRALEAPSGSNVWRNPDMNGHYTPNKTY-----RNSTGQYCREYTOTVVIQKQO 124
OY 142 EIVGTACRQPD 152
DB 125 KAYGNACRQPD 135

RESULT 7
O9K4W8 PRELIMINARY; PRT; 144 AA.
AC O9K4W8;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)

```

```

DE 17 KDA ANTIGEN (FRAGMENT).
OS Male-killing Rickettsia from Adalia decempunctata.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_Taxid=120393;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20575219; Pubmed=11133455;
RA Schultenburg H.G.V.D., Habig M., Sloggett J.J., Webberley M.K.,
RA Bertrand D., Hurst G.D., Majerus M.E.N.;
RT "Incidence of male-killing Rickettsia spp. (alpha-Proteobacteria) in
RT the Ten-Spot ladybird Beetle Adalia decempunctata L. (Coleoptera:
RT Coccinellidae).";
RL Appl. Environ. Microbiol. 67:270-277(2001).
DR EMBL: AJ269516; CAB96381.1; -.
FT NON_TER 1
OX NCBI_Taxid=1
SQ SEQUENCE 144 AA; 14801 MW; C825472F16A56AE7 CRC64;

Query Match 31.0%; Score 259.5; DB 2; Length 144;
Best Local Similarity 39.7%; Pred. No. 6.3e-16;
Matches 52; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

OY 26 SROEVGATGAVVGVGAGQLFGKSGRYAMAIGGAVLGILGSKIGOSMPDQDK-----IK 81
DB 17 NKQGTGTLGGAGGALLGSGFGKSGKQGLVGVGALLGAVLGGOIGAGMDQDRRLAELT 76
OY 82 LNQSLKRYKAGQVTRWRNPDTGNSYSVEPVRTYORYNKQERRQOYCREFOQKAMIAQOK 141
DB 77 SQRALEAPSGSNVWRNPDMNGHYTPNKTY-----RNSTGQYCREYTOTVVIQKQO 131
OY 142 EIVGTACRQPD 152
DB 132 KSYGNACRQPD 142

RESULT 8
O53154 PRELIMINARY; PRT; 154 AA.
AC O53154;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMblrel. 08, last annotation update)
DE (CLONE PRB PISF 1), 5' END CDS (FRAGMENT).
OS Rickettsia sp.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_Taxid=789;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93084757; Pubmed=1452660;
RA Baird R.W., Lloyd M., Stenos J., Ross B.C., Stewart R.S., Dwyer B.;
RT "Characterization and comparison of Australian human spotted fever
RT group rickettsiae.";
RL J. Clin. Microbiol. 30:2896-2902(1992).
DR EMBL: M99391; AAA73386.1; -.
FT NON_TER 154
OX NCBI_Taxid=154
SQ SEQUENCE 154 AA; 15849 MW; F5C35855EDB439D2 CRC64;

Query Match 30.1%; Score 252; DB 2; Length 154;
Best Local Similarity 35.5%; Pred. No. 3.3e-15;
Matches 55; Conservative 29; Mismatches 57; Indels 14; Gaps 4;

OY 9 SLLIITIV--FLVGC--AQNFSRQEVGAATGAVGVAGQLFGKSGRYAMAIGGAVLG 63
DB 5 SKIMIALATSMLOACNPGCMNKGCTLLGGAGGALLGSGFGKQGLVGVGALLG 64
OY 64 GLISKIGOSMPDQDK-----IKLNQSLKRYKAGQVTRWRNPDTGNSYSVEPVRTYORYNK 119
DB 65 AVIGQIGAGDEQDRRLAELTSQRALEAPSGSNVWRNPDMNGSYGVTPNKTYRNSTG 124
OY 120 QERRQOYCREFOQKAMIAQOKOEIYGTACRQPDGR 154

```

Db 125 OD-----CRYVTVTVIGKQKAYGNACRPPDQ 154

RESULT 9

09F909 PRELIMINARY; PRT; 151 AA.
 AC 09F909;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DE OUTER MEMBRANE PROTEIN (FRAGMENT).
 OS Rickettsia helvetica.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OX Rickettsiaceae; Rickettsiae; Rickettsia.
 RN NCBI_TaxID=35789;
 RP SEQUENCE FROM N.A.
 RA Nilsson K., Pahlson C.;
 RT "Novel peptide diagnostic reagent and kit for detection of
 Rickettsiosis."
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF181036; FAG09427.1;
 FT NON_TER 151
 SO SEQUENCE 151 AA; 15621 MW; B77407B9C71E4B39 CRC64;

Query Match
 Best Local Similarity 29.5%; Score 247; DB 2; Length 151;
 Matches 55; Conservative 26; Mismatches 57; Indels 14; Gaps 4;

QY 9 SLLIIV---FLVGC--AQNFSROEVGATGAVGAGVGLFGKSGRVAMAIGAVIG 63
 Db 5 SKIMILIALASMLQACNGPGNNKGGITLLGAGAGALLGSQFGKGGQLVGAVGALLG 64
 QY 64 GLISKTIGQSMDDQDK---IKLNSLEKRVKAGQVTRMRNPDGTGSSVSEVRYQYK 119
 Db 65 AVIGQIVAGMDQDRRLAEITLSQRLAEAPSGSVNWRNPDNGNYVTPNKTY----- 119
 QY 120 QERROGCRFEOCKAMIAQOKQETIGTACROP 151
 Db 120 RNSTGYCREYTVTVIGKQKAYGNACRPP 151

RESULT 10

09F001 PRELIMINARY; PRT; 131 AA.
 AC 09F001;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DE 17 KDA PROTEIN (FRAGMENT).
 OS Rickettsia sp. California 2.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OX Rickettsiaceae; Rickettsiae; Rickettsia.
 RN NCBI_TaxID=147259;
 RP SEQUENCE FROM N.A.
 RA STRAIN-CALIFORNIA 2;
 RT Raoult D.;
 RL "A new SFG rickettsia isolated from fleas."
 RN Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RA STRAIN-CALIFORNIA 2;
 RT Roux V., Raoult D.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF210693; AAG48554.1;
 FT NON_TER 131
 SO SEQUENCE 131 AA; 13374 MW; 23C8819B29FF860 CRC64;

Query Match
 Best Local Similarity 28.4%; Score 237.5; DB 2; Length 131;
 Matches 48; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

QY 26 SROEVGATGAVGAGVAGVGLFGKSGRVAMAIGAVIGLIGSKTIGQSMDDQDK----IK 81
 Db 10 NKQGTGTLGGAGAGALLGSQFGKGGQLVGAVGALLGAVIGAGMDQDRRLAEIT 69
 QY 82 LNSLEKRVKAGQVTRMRNPDGTGSSVSEVRYQYKORROQYCRFEOCKAMIAQOK 141
 Db 70 SORLEAPSGSVNWRNPDNGNHGYVTPNKTY-----RNSTGYCREYTVTVIGKQ 124
 QY 142 EITGTAC 148
 Db 125 KAYGNAC 131

RESULT 11

052637 PRELIMINARY; PRT; 131 AA.
 AC 052637;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DE 17 KDA ANTIGEN (FRAGMENT).
 OS Rickettsia sp.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OX Rickettsiaceae; Rickettsiae; Rickettsia.
 RN NCBI_TaxID=789;
 RP SEQUENCE FROM N.A.
 RA MEDLINE=94117373; PubMed=8288533;
 RA Warren J.H., Hurst G.D., Zhang W., Breeuwer J.A., Stouthamer R.,
 RT "Rickettsial relative associated with male killing in the ladybird
 beetle (Adalia bipunctata)."
 RL J. Bacteriol. 176:388-394(1994).
 DR EMBL: U04162; AAI19233.1;
 FT NON_TER 131
 SO SEQUENCE 131 AA; 13344 MW; A1DCF71050DF52DF CRC64;

Query Match
 Best Local Similarity 28.4%; Score 237.5; DB 2; Length 131;
 Matches 48; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

QY 26 SROEVGATGAVGAGVAGVGLFGKSGRVAMAIGAVIGLIGSKTIGQSMDDQDK----IK 81
 Db 10 NKQGTGTLGGAGAGALLGSQFGKGGQLVGAVGALLGAVIGAGMDQDRRLAEIT 69
 QY 82 LNSLEKRVKAGQVTRMRNPDGTGSSVSEVRYQYKORROQYCRFEOCKAMIAQOK 141
 Db 70 SORLEAPSGSVNWRNPDNGNHGYVTPNKTY-----RNSTGYCREYTVTVIGKQ 124
 QY 142 EITGTAC 148
 Db 125 KAYGNAC 131

RESULT 12

09L522 PRELIMINARY; PRT; 131 AA.
 AC 09L522;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE 17 KDA SURFACE ANTIGEN (FRAGMENT).
 OS Rickettsia peacockii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OX Rickettsiaceae; Rickettsiae; Rickettsia.
 RN NCBI_TaxID=47589;
 RP SEQUENCE FROM N.A.
 RA STRAIN-DAE100R;
 RX MEDLINE=21091941; PubMed=11157215;
 RX Sinsler J.A., Palmer A.T., Munderloh U.G., Kurtli T.J.;

RT "Isolation of a spotted fever group rickettsia, Rickettsia peacockii,
in a Rocky Mountain wood tick, Dermacentor andersoni, cell line.";
RT Appl. Environ. Microbiol. 67:546-552(2001).
DR EMBL: AF260571; AAF69012.1; -.

FT NON_TER 131 131
SO SEQUENCE 131 AA; 13413 MW; 228C020550CA9D0 CRC64;

Query Match 28.3%; Score 236.5; DB 2; Length 131;
Best Local Similarity 37.8%; Pred. No. 6.6e-14;
Matches 48; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

OY 26 SROEVGAATGAVGVGVGOLFSGKSGRVAMAGAVLGLIGSKIGOSMDQDK----IK 81
DB 10 NKGTGTLLGAGAGALLGSGFGKSGKGLVGVGALLGAVLGGGIGAGMDQRRRLAELT 69
OY 82 LNSGLEKVAQGVTRWNPDTGNSVPEVRYORYNKORROOYCREFOQKAMIAQOK 141
DB 70 SGRALFAPSGSNVEMNPNNGVYTPMKTY-----RNSTGQYCREYQTVYVIGSKQ 124
OY 142 EITGTAC 148
DB 125 KAYGNAC 131

RESULT 13
OJ1208 PRELIMINARY; PRT; 105 AA.

ID OJ1208 PRELIMINARY; PRT; 105 AA.
AC OJ1208;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JUN-1998 (Tremblrel. 05, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE 17 KDA ANTIGEN (FRAGMENT).
OS Rickettsia sp. 'La Copita'.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=94475;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LA COPITA;
RX MEDLINE=98367252; PubMed=9701930;
RA Billings A.N., Yu X.J., Teel P.D., Walker D.H.;
RT "Detection of a spotted fever group rickettsia in Amblyomma cajennense
(Acari: Ixodidae) in south Texas";
RL J. Med. Entomol. 35:474-478(1998).
DR EMBL: AF033499; AAB86943.1; -.
FT NON_TER 105 105
SO SEQUENCE 105 AA; 11236 MW; 7BDD35104701D4FB CRC64;

Query Match 26.7%; Score 223.5; DB 2; Length 105;
Best Local Similarity 41.2%; Pred. No. 7.4e-13;
Matches 47; Conservative 17; Mismatches 33; Indels 17; Gaps 3;

OY 43 GOLFSGSGRVAMAGAVLGLIGSKIGOSMDQDK----IKLNSLEKVAQGVTRW 98
DB 3 GOLVGVG-----GALGAVLGGGIGAGMDQRRRLAELTSGRALFAPSGSNVEMR 54
OY 99 NPDGNSVPEVRYORYNKORROOYCREFOQKAMIAQOKYEITGTACRQPD 152
DB 55 NPNGNNGVYTPMKTY-----RNSTGQYCREYQTVYVIGSKQKAYGNACRQPD 103

RESULT 14
OJ985G4 PRELIMINARY; PRT; 199 AA.
AC OJ985G4;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE MLR7687 PROTEIN.
GN Rhlzobium loti (Mesorhizobium loti).
OS

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Mesorhizobium.
OX NCBI_TaxID=381;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsuno M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP003012; BAB54098.1; -.
KW Complete proteome.
SO SEQUENCE 199 AA; 20517 MW; 986EF20A734637AA CRC64;

Query Match 17.1%; Score 143; DB 16; Length 199;
Best Local Similarity 31.8%; Pred. No. 2.9e-05;
Matches 35; Conservative 15; Mismatches 50; Indels 10; Gaps 3;

OY 50 SGRVAMAGAVLGLIGSKIGOSMDQDKIKL-----NLSLEKVAQGVTRWNPDTGNS 105
DB 92 SKVTKTILITAMDGGLIGSIGNGLSDEKRSALAEAYKALEYTTSCOKYAMKDOASHY 151
OY 106 YSVEPVRYORYNKORROOYCREFOQKAMIAQOKYEITGTACRQPDGRW 155
DB 152 GEVVPAPV-RVGSQD-----CRQYQTVFTGAGVARGTACRNPADGSW 195

RESULT 15
OJ9AGC7 PRELIMINARY; PRT; 77 AA.

ID OJ9AGC7 PRELIMINARY; PRT; 77 AA.
AC OJ9AGC7;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE 17 KDA ANTIGEN (FRAGMENT).
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=785;

RN [1]
RP SEQUENCE FROM N.A.
RA Paspaliaris V., Liedtke B., Vilella L., Whiting J.L.;
RT "Rickettsia typhus and Rickettsia felis rickettsioses in Australia";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF336794; AAK21272.1; -.

FT NON_TER 77 77
SO SEQUENCE 77 AA; 7955 MW; B1E447C037263918 CRC64;

Query Match 16.7%; Score 140; DB 2; Length 77;
Best Local Similarity 38.7%; Pred. No. 1.6e-05;
Matches 29; Conservative 13; Mismatches 29; Indels 4; Gaps 1;

OY 43 GOLFSGSGRVAMAGAVLGLIGSKIGOSMDQDK----IKLNSLEKVAQGVTRW 98
DB 3 GSGFGHGKGLGVGVGALLGAVLGGGIGAGMDQRRRLAELTSGRALFAPSGSNVEMR 62
OY 99 NPDGNSVPEVRYORYNKORROOYCREFOQKAMIAQOKYEITGTACRQPD 113
DB 63 NPNGNNGVYTPMKTY-----RNSTGQYCREYQTVYVIGSKQKAYGNACRQPD 77

Search completed: October 27, 2002, 10:59:52
Job time : 22.1865 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:55:57 : Search time 5.87565 Seconds
(without alignments)
1067.553 Million cell updates/sec

Title: US-09-677-374-2

Perfect score: 836
Sequence: 1 MNRGCLQGSSLIITISVFLWG.....IYGTACRQPDGRMQRVISTEK 162

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	296	35.4	159	17KD_RICPR	P16624 rickettsia
2	294	35.2	159	17KD_RICJA	Q52764 rickettsia
3	293	35.0	159	17KD_RICCN	P05372 rickettsia
4	284.5	34.0	159	17KD_RICTY	P22882 rickettsia
5	276	33.0	154	17KD_RICAU	P60928 rickettsia
6	263	31.5	154	17KD_RICAM	P50927 rickettsia
7	261	31.2	154	17KD_RICPA	P50930 rickettsia
8	261	31.2	154	17KD_RICRH	P50931 rickettsia
9	258	30.9	154	17KD_RICMO	P50929 rickettsia
10	157	18.8	80	17KD_RICCA	P29687 rickettsia
11	113.5	13.6	155	PCP_YEREN	P1484 yersinia en
12	105.5	12.6	155	SLVB_ECOLI	P55741 escherichia
13	105.5	12.6	155	SLVB_SALTY	Q53549 salmonella
14	102	12.2	155	PCP_HAERIN	P10335 haemophilus
15	99	11.8	179	KICFJ_ECOLI	P37766 escherichia
16	90	10.8	526	KICJ_BOVIN	P06334 bos taurus
17	90	10.8	1332	XKDO_BACSU	P43334 bacillus su
18	89	10.6	1585	YOBQ_BACSU	P45931 bacillus su
19	87.5	10.5	72	OSMB_SALTY	P37723 salmonella
20	86.5	10.3	72	OSMB_ECOLI	P17853 escherichia
21	84.5	10.1	243	CYSH_SALTY	P18833 salmonella
22	84.5	10.1	301	STXG_RAT	Q92188 rattus norv
23	84.5	10.1	431	KRE2_CANAL	Q00310 candida alb
24	84.5	10.1	526	VP5_BTUVI	P33476 bluecangue
25	82	9.8	132	V615_AQVAE	O66867 aquilex aeo
26	82	9.8	263	CANS_BOVIN	P13135 bos taurus
27	81.5	9.7	541	NU57_YEAST	P18837 saccharomyc
28	81.5	9.7	747	SPD1_MERCL	P19837 nephila cia
29	80.5	9.6	467	HEM1_MYCLE	P46724 mycobacteri
30	79.5	9.5	113	TKR3_CABEL	P34309 cryocorhabdi
31	79.5	9.5	266	CANS_RABIT	P06813 onycholagus
32	79.5	9.5	359	ATPA_BOVIN	P19482 bos taurus
33	79.5	9.5	543	ATPA_RAT	P15999 rattus norv

ALIGNMENTS

34	79.5	9.5	553	1	ATPA_HUMAN	P25705 homo sapien
35	79	9.4	593	1	KICJ_HUMAN	P13645 homo sapien
36	78.5	9.4	553	1	ATPO_BOVIN	P19483 bos taurus
37	78.5	9.4	553	1	ATPA_MOUSE	Q03263 mus musculu
38	78	9.3	865	1	VGLB_HSYMD	P18538 marek's dis
39	77.5	9.3	219	1	VIAD_ECOLI	P37665 escherichia
40	77	9.2	269	1	CANS_MOUSE	O88456 mus musculu
41	77	9.2	393	1	CSP_PLABR	P14593 plasmodium
42	77	9.2	429	1	CSP_PLAMA	P13815 plasmodium
43	77	9.2	506	1	ATPA_ANASP	P12405 anabena sp
44	77	9.2	514	1	ATPA_THIFE	P4167 thiodacillu
45	76.5	9.2	243	1	CYSH_ECOLI	P17854 escherichia

RESULT 1	17KD_RICPR	STANDARD:	PRT:	159 AA.
AC	P16624;			
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-AUG-1991 (Rel. 19, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	17 kda surface antigen precursor.			
GN	OMP OR RP833.			
OS	Rickettsia prowazekii.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;			
OC	Rickettsiaceae; Rickettsiae; Rickettsia.			
OX	NCBI_Taxid=782;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MADRID E.			
RX	MEDLINE=99039499; PubMed=9823893;			
RA	Medline=S.G.E., Zomrodipour A., Andersson J.O.,			
RA	Anderson B.E., Tzianabos T.;			
RT	"Comparative sequence analysis of a genus-common rickettsial antigen			
RT	gene.";			
RL	J. Bacteriol. 171:5199-5201(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MADRID E.			
RX	MEDLINE=99039499; PubMed=9823893;			
RA	Medline=S.G.E., Zomrodipour A., Andersson J.O.,			
RA	Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,			
RT	Eriksson A.-S., Winkler H.H., Kurland C.G.;			
RT	"The genome sequence of Rickettsia prowazekii and the origin of			
RT	mitochondria.";			
RL	Nature 396:133-140(1998).			
CC	-1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid			
CC	anchor (Probable).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M28482; AAA26378.1; ALT_SEQ.			
DR	EMBL: AJ235273; CAA15258.1; -.			
DR	PIR: D33971; D33971.			
DR	PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.			
KW	Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.			
FT	SIGNAL	1	19	
FT	CHAIN	20	159	17 KDA SURFACE ANTIGEN.
FT	LIPID	20	20	N-ACYL DIGLYCERIDE (PROBABLE).
SQ	SEQUENCE	159 AA.	16672 MM.	A33D404B5EBB071.CRC64;

Query Match	35.4%	Score 296;	DB 1;	Length 159;
Best Local Similarity	38.1%	Pred. No. 9.9e-19;		
Matches	61;	Conservative	30;	Mismatches 55; Indels 14; Gaps 4;

OY 9 SLLIIVS---FLVGC--AONFSROEYGAATGAVVAGVAGOLFCKGSGRVAMALGAVIG 63
 Db 5 SKIMITALASMLACAGNGOSGMKOGTGLLGAGAGALLGSGFQGGKGLVGVGALLG 64
 OY 64 GLIGSKIGOSMDQODK-----IKLNOSLEKVAAGVOTWRNDPTGNSVSEPRITQRYNK 119
 Db 65 AVLGQIGAGMDEODRLRLLETSORALTSAPSGSNIEMRPNPDNGNNGYTPNKTY----- 119
 OY 120 QERROQYCRFPOOKAMIAQOKOIEYTAGCRPDGRMOVIS 159
 Db 120 RNSAGQYCREYTOTVYIGKQOKKTYGNACRPPDGMQOVVN 159

RESULT 2

17KD_RICJA
 ID 17KD_RICJA STANDARD; PRT: 159 AA.
 AC 032764;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 17 kDa surface antigen precursor.
 GN OMP.
 OS Rickettsia japonica.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OX Rickettsiaceae; Rickettsiae; Rickettsia.
 RN NCBI_TaxID=35790;
 RP SEQUENCE FROM N.A.
 RC STRAIN-YH.
 RX MEDLINE=95229950; PubMed=7714214;
 RA Furuya Y., Katayama T., Yoshida Y., Kaiho I.;
 RT "Specific amplification of Rickettsia japonica DNA from clinical
 RU specimens by PCR."
 RL J. Clin. Microbiol. 33:487-489(1995).
 CC -i- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (Probable).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D16515; BAA0365.1;
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1;
 KW Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 159
 FT LIPID 20 20
 SO SEQUENCE 159 AA; 16554 MW; CDCE/CEHDCD6B41 CRC64;

Query Match

Best Local Similarity 35.2%; Score 294; DB 1; Length 159;
 Matches 61; Conservative 30; Mismatches 55; Indels 14; Gaps 4;
 OY 9 SLLIIVS---FLVGC--AONFSROEYGAATGAVVAGVAGOLFCKGSGRVAMALGAVIG 63
 Db 5 SKIMITALASMLACAGNGOSGMKOGTGLLGAGAGALLGSGFQGGKGLVGVGALLG 64
 OY 64 GLIGSKIGOSMDQODK-----IKLNOSLEKVAAGVOTWRNDPTGNSVSEPRITQRYNK 119
 Db 65 AVLGQIGAGMDEODRLRLLETSORALTSAPSGSNIEMRPNPDNGNNGYTPNKTY----- 119
 OY 120 QERROQYCRFPOOKAMIAQOKOIEYTAGCRPDGRMOVIS 159
 Db 120 RNSAGQYCREYTOTVYIGKQOKKTYGNACRPPDGMQOVVN 159

RESULT 3

17KD_RICCN
 ID 17KD_RICCN STANDARD; PRT: 159 AA.

AC P05372;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1990 (Rel. 15, Last sequence update)
 DE 17 kDa surface antigen precursor.
 GN OMP OR RC1287.
 OS Rickettsia conorii, and
 OC Bacteria; Proteobacteria;
 OX Rickettsiaceae; Rickettsiae; Rickettsia.
 RN NCBI_TaxID=781, 783;
 RP SEQUENCE FROM N.A.
 RC SPECIES=R.conorii, and R.rickettsii;
 RX MEDLINE=89359171; PubMed=2768201;
 RA Anderson B.E., Tzianabos T.;
 RT "Comparative sequence analysis of a genus-common rickettsial antigen
 RU gene."
 RL J. Bacteriol. 171:5199-5201(1989).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC SPECIES=R.conorii.
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samsou D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii."
 RL Science 293:2093-2098(2001).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC SPECIES=R.rickettsii;
 RX MEDLINE=8722152; PubMed=3108222;
 RA Anderson B.E., Regnery R.L., Carlone G.M., Tzianabos T., McDeade J.E.,
 RA Fu Z.Y., Bellini W.J.;
 RT "Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia
 RU rickettsii."
 RL J. Bacteriol. 169:2385-2390(1987).
 RN (4)
 RP SEQUENCE OF 1-30 FROM N.A.
 RC SPECIES=R.rickettsii;
 RX MEDLINE=89008059; PubMed=3139629;
 RA Anderson B.E., Baumstark B.R., Bellini W.J.;
 RT Expression of the gene encoding the 17-kilodalton antigen from
 RU Rickettsia rickettsii: transcription and posttranslational
 RL modification."
 RL J. Bacteriol. 170:4493-4500(1988).
 CC -i- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (Probable).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M28479; AAA26379.1;
 DR EMBL: M28480; AAA26376.1;
 DR EMBL: AF008675; AAL03825.1;
 DR EMBL: M16486; AAA26381.1;
 DR EMBL: J03371; -; NOT_ANNOTATED_CDS.
 DR PIR: A25972; A25972.
 DR PIR: A31836; A31836.
 DR PIR: A33971; A33971.
 DR PIR: B33971; B33971.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1;
 KW Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.
 FT SIGNAL 1 19
 FT CHAIN 20 159
 FT LIPID 20 20
 FT LIPID 20 20
 FT CONFLICT 146 146
 FT CONFLICT 153 153
 FT CONFLICT 153 153
 G -> E (IN REF. 3)
 G -> E (IN REF. 3)

SQ SEQUENCE 159 AA; 16581 MW; 206A2BBF74FCE169 CRC64;

Query Match 35.0%; Score 293; DB 1; Length 159;

Best Local Similarity 38.1%; Pred. No. 1 Be-18; Matches 61; Conservative 29; Mismatches 56; Indels 14; Gaps 4;

OY 9 SLLIIISV---FLVGC--AONFSROEVGAATGAVGVACQLFGKSGRYAMAIIGAVLG 63
 AC P22882; 17KD_RICAM STANDARD; PRT; 159 AA.
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 17 kDa surface antigen precursor.
 GN OMP.
 OS Rickettsia typhi.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_TaxID=785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89359171; PubMed=2768201;
 RA Anderson B.E., Tzianabos T.;
 RT "Comparative sequence analysis of a genus-common rickettsial antigen gene."
 RT J. Bacteriol. 171:5199-5201(1989).
 RL -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: M28481; AAA26377.1; -
 DR PIR: C33971; C33971.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
 FT LIPID 20 20 N-ACYL DIGLICERIDE (PROBABLE).
 SO SEQUENCE 159 AA; 16549 MW; 08973E264E8D8C8 CRC64;

Query Match 34.0%; Score 284.5; DB 1; Length 159;

Best Local Similarity 39.9%; Pred. No. 9.5e-18; Matches 55; Conservative 25; Mismatches 49; Indels 9; Gaps 2;

OY 26 SREVGAAAGVAVGVAAGLFGKSGRYAMAIIGAVLGIGKSGMDQDK-----IK 81
 AC P50927; 17KD_RICAM STANDARD; PRT; 154 AA.
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 17 kDa surface antigen precursor (fragment).
 GN OMP.
 OS Rickettsia amblyomii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_TaxID=33989;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=MO 85-1084;
 SO SEQUENCE 154 AA; 16549 MW; 08973E264E8D8C8 CRC64;

DB 142 TTYGNACRQPDGOMOVN 159

RESULT 5

ID 17KD_RICAM STANDARD; PRT; 154 AA.
 AC P50927;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 17 kDa surface antigen precursor (fragment).
 GN OMP.
 OS Rickettsia australis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_TaxID=787;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Baird R.W., Ross B., Dwyer B.;
 RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: M74042; AAA26394.1; -
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
 FT LIPID 20 20 N-ACYL DIGLICERIDE (PROBABLE).
 FT NON_TER 154 154
 SO SEQUENCE 154 AA; 15967 MW; E3AA833346FAC320 CRC64;

Query Match 33.0%; Score 276; DB 1; Length 154;

Best Local Similarity 38.1%; Pred. No. 4.9e-17; Matches 59; Conservative 27; Mismatches 55; Indels 14; Gaps 4;

OY 9 SLLIIISV---FLVGCNQ--NFSROEVGAATGAVGVACQLFGKSGRYAMAIIGAVLG 63
 AC P50927; 17KD_RICAM STANDARD; PRT; 154 AA.
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 17 kDa surface antigen precursor (fragment).
 GN OMP.
 OS Rickettsia amblyomii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_TaxID=33989;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=MO 85-1084;
 SO SEQUENCE 154 AA; 16549 MW; 08973E264E8D8C8 CRC64;

Query Match 33.0%; Score 276; DB 1; Length 154;

Best Local Similarity 38.1%; Pred. No. 4.9e-17; Matches 59; Conservative 27; Mismatches 55; Indels 14; Gaps 4;

OY 9 SLLIIISV---FLVGCNQ--NFSROEVGAATGAVGVACQLFGKSGRYAMAIIGAVLG 63
 AC P50927; 17KD_RICAM STANDARD; PRT; 154 AA.
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 17 kDa surface antigen precursor (fragment).
 GN OMP.
 OS Rickettsia amblyomii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_TaxID=33989;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=MO 85-1084;
 SO SEQUENCE 154 AA; 16549 MW; 08973E264E8D8C8 CRC64;

RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
 RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U11013; AAB07704.1;
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
 FT LIPID 20 20
 FT NON_TER 154 154 N-ACYL DIGLYCERIDE (PROBABLE).
 SQ SEQUENCE 154 AA; 15879 MW; E4FBEC29D943581 CRC64;

Query Match
 Best Local Similarity 31.5%; Score 263; DB 1; Length 154;
 Matches 57; Conservative 27; Mismatches 57; Indels 14; Gaps 4;

QY 9 SLLIISV---FLVGC--AONFSROEVAATGAVVAGVAGOLFSGSGRVAMAGAVLG 63
 DB 5 SKIMITALAASMLQACNGPGMKNKGOTGLLGAGGALLGSGFGKGLVGVGVALLG 64
 QY 64 GLIGSKIGQSMDOOK-----IKLNSLEKVKAGVTRRRNDPTGNSYSVEPRTYQRYNK 119
 DB 65 AVLGQIGAGMDEDDRIELTSQALETAPSGSNVEMRNPDNGNGYTPNKTY----- 119
 QY 120 QERRQOYCREFOQKAMTAGOKOETIYTACROPDGR 154
 DB 120 RNSTGQYCREYTOTVIVIGKQOKAYGNACLOPDGQ 154

RESULT 7
 17KD_RICPA STANDARD; PRT; 154 AA.
 AC P50931;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 17 kDa surface antigen precursor (Fragment).
 GN OMP.
 OS Rickettsia parkeri.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 RN NCBI_TaxID=35792;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MACULATUM;
 RA Pretzman C.I., Stothard D.R., Ralph D., Clark J.B., Fuerst P.A.;
 RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U17008; AAA82040.1;
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.

FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
 FT NON_TER 154 154
 SQ SEQUENCE 154 AA; 15897 MW; 5D06F45FDDBD5EEC CRC64;

Query Match
 Best Local Similarity 31.2%; Score 261; DB 1; Length 154;
 Matches 56; Conservative 28; Mismatches 57; Indels 14; Gaps 4;

QY 9 SLLIISV---FLVGC--AONFSROEVAATGAVVAGVAGOLFSGSGRVAMAGAVLG 63
 DB 5 SKIMITALAASMLQACNGPGMKNKGOTGLLGAGGALLGSGFGKGLVGVGVALLG 64
 QY 64 GLIGSKIGQSMDOOK-----IKLNSLEKVKAGVTRRRNDPTGNSYSVEPRTYQRYNK 119
 DB 65 AVLGQIGAGMDEDDRIELTSQALETAPSGSNVEMRNPDNGNGYTPNKTY----- 119
 QY 120 QERRQOYCREFOQKAMTAGOKOETIYTACROPDGR 154
 DB 120 RNSTGQYCREYTOTVIVIGKQOKAYGNACLOPDGQ 154

RESULT 8
 17KD_RICRH STANDARD; PRT; 154 AA.
 AC P50931;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 17 kDa surface antigen precursor (Fragment).
 GN OMP.
 OS Rickettsia rhipicephali.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 RN NCBI_TaxID=33992;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
 RL Submitted (AUG-1994) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U11020; AAB07706.1;
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
 FT LIPID 20 20
 FT NON_TER 154 154 N-ACYL DIGLYCERIDE (PROBABLE).
 SQ SEQUENCE 154 AA; 15895 MW; 0CF85AD5D96DEFFB CRC64;

Query Match
 Best Local Similarity 31.2%; Score 261; DB 1; Length 154;
 Matches 56; Conservative 28; Mismatches 57; Indels 14; Gaps 4;

QY 9 SLLIISV---FLVGC--AONFSROEVAATGAVVAGVAGOLFSGSGRVAMAGAVLG 63
 DB 5 SKIMITALAASMLQACNGPGMKNKGOTGLLGAGGALLGSGFGKGLVGVGVALLG 64
 QY 64 GLIGSKIGQSMDOOK-----IKLNSLEKVKAGVTRRRNDPTGNSYSVEPRTYQRYNK 119
 DB 65 AVLGQIGAGMDEDDRIELTSQALETAPSGSNVEMRNPDNGNGYTPNKTY----- 119
 QY 120 QERRQOYCREFOQKAMTAGOKOETIYTACROPDGR 154
 DB 120 RNSTGQYCREYTOTVIVIGKQOKAYGNACLOPDGQ 154


```

RESULT 9
17KD_RICMO STANDARD; PRT; 154 AA.
ID 17KD_RICMO
AC P50929;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia montana.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=33991;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OHIO 83-441;
RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
CC EMBL: U11017; AAB07705.1; -
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KM Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15881 MW; A09C53B8769E31DA CRC64;

Query Match 30.9%; Score 258; DB 1; Length 154;
Best Local Similarity 36.1%; Pred. No. 1.7e-15;
Matches 56; Conservative 28; Mismatches 57; Indels 14; Gaps 4;

QY 9 SLLIIVS---FLVGC---AQNFSRQEVGAATGAVGVAGQLPEKSGSRVAMAGVAVL 63
DB 5 SKMIITIALASMLQACPGCGMKNQGTGLLGAGALLSOFQGGQLVGVGVALLG 64
QY 64 GLTSGKIGQSMDOODK-----IKLNQSLKRYKAGOVTRMRNPDGNSVSEPVRTYQRYNK 119
DB 65 AVLOGQIGAGMDEQDRRLALETISQRALETAPSGSNVEMRNPDNGNYGVTPNKTY----- 119
QY 120 QERRQOYCREFOQKAMITAGQKQEIYGTACROPDGR 154
DB 120 RNSTGYCREYQTVVIGKQKQKAYGNACLOPDQO 154

RESULT 10
17KD_RICCA STANDARD; PRT; 80 AA.
ID 17KD_RICCA
AC P29697;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen (Fragment).
GN OMP.
OS Rickettsia canada.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=788;
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE=92108069; Pubmed=1729713;

```

```

RA Azad A.F., Sacchi J.B. Jr., Nelson W.M., Dasch G.A.,
RA Schmidtmann E.T., Carl M.;
RT "Genetic characterization and transovarial transmission of a
RT typhus-like rickettsia found in cat fleas."
RL Proc. Natl. Acad. Sci. U.S.A. 89:43-46(1992).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
CC EMBL: M82879; -; NOT_ANNOTATED_CDS.
DR InterPro; IPR000437; PROKAR_LIPOPROT.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
KM Outer membrane; Lipoprotein; Antigen.
FT NON_TER 1 1
FT NON_TER 80 80
SQ SEQUENCE 80 AA; 8372 MW; AD289A48EAB19E0E CRC64;

Query Match 18.8%; Score 157; DB 1; Length 80;
Best Local Similarity 40.3%; Pred. No. 4e-07;
Matches 31; Conservative 15; Mismatches 27; Indels 4; Gaps 1;

QY 43 GOLFSGSGRYAMAGVAVLGLGKSGKIGQSMDOODK-----IKLNQSLKRYKAGOVTRMR 98
DB 1 GSGFGKKGGLIGVAGCALLGAILGNQIGAGMDEQDRRLALETISQRALETTPSTSIEMR 60
QY 99 NPDTGNSVSEPVRTYQ 115
DB 61 NPDGNGYGVTPSKTYK 77

RESULT 11
PCP_YEREN STANDARD; PRT; 155 AA.
ID PCP_YEREN
AC P31484;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane lipoprotein pcv precursor.
GN PCP OR PCPY.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51872 / WA-C / SEROTYPE O:8;
RL MEDLINE=92121089; Pubmed=1732192;
RA Baunler A.J., Hanke K.;
RT "A lipoprotein of Yersinia enterocolitica facilitates ferrioxamine
RT uptake in Escherichia coli."
RL J. Bacteriol. 174:1029-1035(1992).
CC -1- FUNCTION: FACILITATES FERRIOXAMINE UPTAKE.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -----
CC -1- SIMILARITY: TO E. COLI AND S. TYPHIMURUM SLXB AND TO
CC H. INFLUENZAE PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
CC EMBL: X60448; CAA42977.1; -

```

```

DR   PIR: S23787; S23787.
DR   HSSP: P00778; 1P04.
DR   PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Outer membrane; Lipoprotein; Signal.
FT   CHAIN          1           POTENTIAL.
FT   LIPID          18           OUTER MEMBRANE LIPOPROTEIN PCP.
FT   LIPID          18           N-ACYL DIGLYCERIDE (POTENTIAL).
SQ   SEQUENCE      155 AA; 15362 MW; 8ADBE2132E849FA CRC64;
Query Match
Best Local Similarity 24.7%; Score 113.5; DB 1; Length 155;
Matches 37; Conservative 19; Mismatches 43; Indels 51; Gaps 4;
OY   10 SLITISVELVCAQN-----FSRQE-----VGAAT 34
DB   7 AVALAAVLTGCAANNNTLSGDVFSASQAKOVQVTVTGLLSVPEVTIOGGDDNNVGAIG 66
OY   35 GAVGVAGQGLCKSGSRVAMAGAVAGLIGSKIGQSMDDQDKIKL-----82
DB   67 GAVLGGLFNTVGGGTGRSLATTAAGAVAGGAGQVAGAMNRTDGVLEVRKDDGTTILV 126
OY   83 --NQSLEKVKAGQVTRMRNPDTGNSVSEP 110
DB   127 VOKQGPTRFSVGO--RVMLASGSGTIVTSP 154

RESULT 12
SLYB_ECOLI
ID   SLYB_ECOLI          STANDARD:          PRT; 155 AA.
AC   P35741; P76183;
DT   01-NOV-1997 (Rel. 35, Created)
DD   16-OCT-2001 (Rel. 35, Last sequence update)
DE   Outer membrane lipoprotein slyB precursor.
GN   SLYB OR B1641 OR Z2655 OR ECS2350.
OS   Escherichia coli, and
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ludwig A., Tengcl C., Bauer S., Bubert A., Benz R., Mollenkopf H.-J.,
RA   Goebel W.;
RT   'slyB', a regulatory protein from Salmonella typhimurium, induces a
RT   haemolytic and pore-forming protein in Escherichia coli.;
RL   Mol. Genet. 249:474-486(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   STRAIN-K12 / MG1655;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RT   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   STRAIN-K12;
RA   MEDLINE-97251357; PubMed-9097039;
RA   Itoh T., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA   Kitagawa K., Makino K., Kimura S., Kitakawa M.,
RA   Motomura K., Nakade S., Nakamura Y., Mizobuchi K., Mori H., Mori T.,
RA   Oshima T., Saito N., Sempel G., Seki Y., Sivasubraman S.,
RA   Yamamoto Y., Horuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RT   DNA Res. 3:363-377(1996).

```

```

RP   SEQUENCE FROM N.A.
RC   STRAIN-0157:H7 / EDL933 / ATCC 700927;
RC   MEDLINE-21074935; PubMed-11206551;
RC   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RC   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick K.,
RC   Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RC   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RC   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.",
RT   Nature 409:529-533(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   STRAIN-0157:H7 / RIMD 0509952;
RA   MEDLINE-21156231; PubMed-11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Ikeda T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.",
RT   DNA Res. 8:11-22(2001).
CC   -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC   anchor (Potential).
CC   -1- SIMILARITY: TO S. TYPHIMURIUM SLYB, H. INFLUENZAE PCP AND
CC   Y. ENTEROCOLITICA PCP.
CC   THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC   use by non-profit institutions as long as there are no restrictions on its
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   -- send an email to license@sib-sib.ch).
DR   EMBL: AE000259; AAC74713.1;
DR   EMBL: D90807; BAI15402.1;
DR   EMBL: AE005387; AAG56630.1;
DR   EMBL: AP002558; BAB35773.1;
DR   ECGene; EGI3409; slyB
DR   PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Outer membrane; Lipoprotein; Signal; Complete proteome.
FT   CHAIN          1           POTENTIAL.
FT   LIPID          18           OUTER MEMBRANE LIPOPROTEIN PCP.
FT   LIPID          18           N-ACYL DIGLYCERIDE.
FT   CONFLICT       98          A -> T (IN REF. 2).
SQ   SEQUENCE      155 AA; 15602 MW; 543BB84D069A5FA3 CRC64;
Query Match
Best Local Similarity 12.6%; Score 105.5; DB 1; Length 155;
Matches 33; Conservative 21; Mismatches 42; Indels 49; Gaps 4;
OY   13 IISVELVCAQN-----FSRQE-----VGAATGAV 37
DB   10 MYGLSLVGCYNNNTLSGDVFTASEAKQVNSYGTIVNRPVIOGCDSDNVYIGAV 69
OY   38 VGVAGQGLCKSGSRVAMAGAVAGLIGSKIGQSMDDQDKIKLQSLKVKAGQVTRW 97
DB   70 LGGLFNTVGGGTGRSLATTAAGAVAGGAGQVAGAMNRTDGVLEVRKDDGTTILV 126
OY   98 RNPDTGNSVSEPRTVRYRNKQER 122
DB   119 ---DDGNTIMVVKQGNTRFSRQGR 140

RESULT 13
SLYB_SALTY
ID   SLYB_SALTY          STANDARD:          PRT; 155 AA.
AC   O53549;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DE   Outer membrane lipoprotein slyB precursor

```

```

GN SLVB OR STM1445 OR STY1677.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium;
RX MEDLINE=96133688; PubMed=8544813;
RA Ludwig A., Tengel C., Bauer S., Hubert A., Benz R., Mollenkopf H.-J.,
RA Goebel W.;
RT "SLVA, a regulatory protein from Salmonella typhimurium, induces a
RT hemolytic and pore-forming protein in Escherichia coli.";
RL Mol. Gen. Genet. 249:474-486(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Portolillo S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gea P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar typhi CT18.";
RL Nature 413:848-852(2001).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Potential).
CC -1- SIMILARITY: TO E.COLI SLVB, H.INFLUENZAE PCP AND Y.ENTEROCOLITICA
CC PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S80790; AAB35871.2; -
DR EMBL: AE008762; AAL20367.1; -
DR EMBL: AL627271; CADU1922.1; -
DR StGene; SG10573; slvb.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KM Outer membrane; Lipoprotein; signal; Complete proteome.
FT STGNL 1 17
FT CHAIN 18 155 OUTER MEMBRANE LIPOPROTEIN SLVB.
FT LIPID 18 18 N-ACYL DIGLYCERIDE.
SQ SEQUENCE 155 AA; 15548 MW; 82FDDCDBABD55A7 CRC64;

```

Query Match 12.6%; Score 105.5; DB 1; Length 155;
 Best Local Similarity 24.6%; Pred. No. 0.02; Mismatches 33; Gaps 3;
 Matches 34; Conservative 19; Indels 52;

```

OY 4 GCLQGSGL-----IISVFLVCAQNFSGROEVAATGAAGVAVGAGQ 44
DB 17 GCVNNDLSGSDVYTA$BAKQVQNTYGTIVNRPVQIOGGDDSNVIGALIGAVLGGFLGN 76

```

```

OY 45 LFGKSGRVAMATGCAVLGGLIGSKIGOSMDQDKIKLNSLEKXKAGQYTRMRPDTGN 104
DB 77 TIGGCTGRSLATAGAAVAGAGGQVQ$AMKKTQGVEL--EIRK-----DDGN 122
OY 105 SYSEVPRTYQRYNKR 122
DB 123 TIMVYQKQGNTRFSAGOR 140

```

RESULT 14
 PCP_HAEIN STANDARD; PRT; 155 AA.
 ID PCP_HAEIN
 AC P10325;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DE Outer membrane lipoprotein and an antigenically related 15,000-dalton
 DE cross-reacting lipoprotein.
 DE PCP OR LPP OR H1579.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_Taxid=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86115138; PubMed=2828309;
 RA Deich R.A., Metcalf B.J., Finn C.W., Farley J.E., Green B.A.;
 RT Cloning of genes encoding a 15,000-dalton peptidoglycan-associated
 RT outer membrane lipoprotein and an antigenically related 15,000-dalton
 RT protein from Haemophilus influenzae.";
 RL J. Bacteriol. 170:489-498(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RL Science 269:496-512(1995).
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor.
 CC -1- SIMILARITY: TO E.COLI AND S.TYPHIMURIUM SLVB AND TO
 CC Y.ENTEROCOLITICA PCP.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M18877; AAA24938.1; -
 DR EMBL: U32832; AAC23228.1; -
 DR PIR: B28543; B28543.
 DR TIGR: H11579; -
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KM Outer membrane; Lipoprotein; signal; Complete proteome.
 FT STGNL 1 18
 FT CHAIN 19 155 OUTER MEMBRANE LIPOPROTEIN PCP.
 FT LIPID 19 19 N-ACYL DIGLYCERIDE.
 FT CONFLICT 135 143 CSLVAEFLV -> VAGRRVRI (IN REF. 1).
 SQ SEQUENCE 155 AA; 15425 MW; D7880327FCFC985 CRC64;

Query Match
Best Local Similarity 12.2%; Score 102; DB 1; Length 155;
Matches 21; Conservative 9; Mismatches 23; Indels 0; Gaps 0;

QY 30 VGANTGAVGVAGQLFGKSGSRVMAIGAVLGLSGISQSMQDQKRL 82
DB 62 VGTLAGGAGLGIAGTSTIGGGRQVIAAVGALGALGASKEKMSQVGAEL 114

RESULT 15
YCFJ_ECOLI

ID YCFJ_ECOLI STANDARD; PRT: 179 AA.

AC P37796; P75951; 01-OCT-1994 (Rel. 30, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE 01-MAR-2002 (Rel. 41, Last annotation update)

GN YCFJ OR B1110.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OX NCBI_TaxID=562;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN-K12; MEDLINE-9742617; PubMed-9278503;

RA Blatter F.R., Plunkett G. III, Bloch C.A., Perma N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN (2)

RP SEQUENCE FROM N.A.

RC STRAIN-K12; MEDLINE-97061202; PubMed-8905232;

RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,

RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,

RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,

RA Mori H., Motomura K., Nakamura Y., Nishio Y., Saito N.,

RA Sampaio G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,

RA Yano M., Horiiuchi T.;

RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome

corresponding to the 12.7-28.0 min region on the linkage map.";

RL DNA Res. 3:137-155(1996).

RN (3)

RP SEQUENCE OF 1-63 FROM N.A.

RC STRAIN-K12; MEDLINE-81236546; PubMed-6265208;

RA Young J.G., Rogers B.L., Campbell H.D., Jaworski A., Shaw D.C.;

RT "Nucleotide sequence coding for the respiratory NADH dehydrogenase of

Escherichia coli. UUG initiation codon.";

RL Eur. J. Biochem. 116:165-170(1981).

RN (4)

RP IDENTIFICATION.

RC MEDLINE-95075659; PubMed-7984428;

RA Borodovsky M., Rudd K.E., Koonin E.V.;

RT "Intrinsic and extrinsic approaches for detecting genes in a

bacterial genome.";

RL Nucleic Acids Res. 22:4756-4767(1994).

RN (5)

RP SIMILARITY: TO RICKETTSIA 17 KDa SURFACE ANTIGEN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL: A6000211; AAC74194.1;

CC EMBL: D90746; BAA35925.1;

CC DR EMBL: V00306; NOT_ANNOTATED_CDS.

DR Ecogene; EGI2444; ycfj.
KW Hypothetical protein; Transmembrane; Complete proteome.
SQ TRANSMEM 5 25 POTENTIAL.
SEQUENCE 179 AA; 18920 MW; BA5EB0B56D45609 CRC64;
Query Match
Best Local Similarity 11.8%; Score 99; DB 1; Length 179;
Matches 26; Conservative 12; Mismatches 32; Indels 26; Gaps 1;
QY 31 GATGAVGVAGQLFGKSGSRVMAIGAVLGLSGISQSMQDQKRLKMSLEKVK 90
DB 73 GSVLGAAGVGVGHGFGGGRKMDVATVGAAGGAGVAGNOLGSLQESD----- 120
QY 91 AGQVTRMRNPDGNSYSVEPVPTIORYNKKORROQY 126
DB 121 -----YTTTQQRCKTVYDKSEKMLGY 142
Search completed: October 27, 2002, 10:56:29
Job time: 8.87565 secs

B33971
Rickettsia common antigen precursor - Rickettsia conorii

C:Species: Rickettsia conorii
C:Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 20-Aug-1999
C:Accession: B33971
R:Anderson, B.E.; Tzianabos, T.
J. Bacteriol. 171, 5199-5201, 1989
A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.
A:Reference number: A33971; MUID:89359171
A:Accession: B33971
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-159 <AND>
A:Cross-references: GB:M28479; NID:g152463; PIDN:AAA26379.1; PID:g152464
C:Superfamily: rickettsial common antigen

Query Match
Best Local Similarity 35.0%; Score 293; DB 2; Length 159;
Matches 61; Conservative 29; Mismatches 56; Indels 14; Gaps 4;
OY 9 SLLIISV---FLVGC--AONFSRQEVGAATGAVGVAGVAGQLFGKSGRVAMAIGAVLG 63
D 5 SKIMIALATSMLOACNGPGMKNKGTGTLTGAGAGALLGSQFGKKGQLVGVGALLG 64
OY 64 GLISKIGQSMDOODK----IKLNSLEKYKAGQVTRMRNPDTGNSYSVEPVRYQRYNK 119
D 65 AVLGGQIGAGMDQDRRLAELTSGRALETAPSGSNVEMRNPDNGNGVYTPNKTY----- 119
OY 120 QERRQYCRFQOKAMIAQOGKEITGTACROPDGRMOVIS 159
D 120 RNSTGYCRREYTOTVYIGKQOKAYGNACRQPDGMQVYN 159

RESULT 3
A33971
Rickettsial common antigen precursor - Rickettsia rickettsii
C:Species: Rickettsia rickettsii
C:Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 20-Aug-1999
R:Anderson, B.E.; Tzianabos, T.
J. Bacteriol. 171, 5199-5201, 1989
A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.
A:Reference number: A33971; MUID:89359171
A:Accession: A33971
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-159 <AND>
A:Cross-references: GB:M28480; NID:g152457; PIDN:AAA26376.1; PID:g152458
C:Superfamily: rickettsial common antigen

Query Match
Best Local Similarity 35.0%; Score 293; DB 2; Length 159;
Matches 61; Conservative 29; Mismatches 56; Indels 14; Gaps 4;
OY 9 SLLIISV---FLVGC--AONFSRQEVGAATGAVGVAGVAGQLFGKSGRVAMAIGAVLG 63
D 5 SKIMIALATSMLOACNGPGMKNKGTGTLTGAGAGALLGSQFGKKGQLVGVGALLG 64
OY 64 GLISKIGQSMDOODK----IKLNSLEKYKAGQVTRMRNPDTGNSYSVEPVRYQRYNK 119
D 65 AVLGGQIGAGMDQDRRLAELTSGRALETAPSGSNVEMRNPDNGNGVYTPNKTY----- 119
OY 120 QERRQYCRFQOKAMIAQOGKEITGTACROPDGRMOVIS 159
D 120 RNSTGYCRREYTOTVYIGKQOKAYGNACRQPDGMQVYN 159

RESULT 4
G97860
17k surface antigen precursor [Imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: G97860
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barde, V.; Samson, D.; R
Science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: G97860
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-159 <KUR>
A:Cross-references: GB:A6006914; PIDN:AAL03825.1; PID:g15620425; GSPDB:GN00173
A:Gene: omp
C:Superfamily: rickettsial common antigen

Query Match
Best Local Similarity 35.0%; Score 293; DB 2; Length 159;
Matches 61; Conservative 29; Mismatches 56; Indels 14; Gaps 4;
OY 9 SLLIISV---FLVGC--AONFSRQEVGAATGAVGVAGVAGQLFGKSGRVAMAIGAVLG 63
D 5 SKIMIALATSMLOACNGPGMKNKGTGTLTGAGAGALLGSQFGKKGQLVGVGALLG 64
OY 64 GLISKIGQSMDOODK----IKLNSLEKYKAGQVTRMRNPDTGNSYSVEPVRYQRYNK 119
D 65 AVLGGQIGAGMDQDRRLAELTSGRALETAPSGSNVEMRNPDNGNGVYTPNKTY----- 119
OY 120 QERRQYCRFQOKAMIAQOGKEITGTACROPDGRMOVIS 159
D 120 RNSTGYCRREYTOTVYIGKQOKAYGNACRQPDGMQVYN 159

RESULT 5
C33971
Rickettsial common antigen precursor - Rickettsia typhi
C:Species: Rickettsia typhi
C:Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 20-Aug-1999
R:Anderson, B.E.; Tzianabos, T.
J. Bacteriol. 171, 5199-5201, 1989
A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.
A:Reference number: A33971; MUID:89359171
A:Accession: C33971
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-159 <AND>
A:Cross-references: GB:M28481; NID:g152459; PIDN:AAA26377.1; PID:g152460
C:Superfamily: rickettsial common antigen

Query Match
Best Local Similarity 34.0%; Score 284.5; DB 2; Length 159;
Matches 55; Conservative 25; Mismatches 49; Indels 9; Gaps 2;
OY 26 SRQEVGAATGAVGVAGVAGQLFGKSGRVAMAIGAVLGGLISKIGQSMDOODK---IK 81
D 27 NKQGTGTLTGAGAGALLGSQFGKKGQLVGVGALLGASLDEDRRLLELT 86
OY 82 LNSLEKYKAGQVTRMRNPDTGNSYSVEPVRYQRYNKQERRQYCRFQOKAMIAQOGK 141
D 87 SQRLAESAPSGSNVEMRNPDNGNGVYTPNKTY-----RNSTGYCRREYTOTVYIGKQ 141
OY 142 EIVGTACROPDGRMOVIS 159
D 142 TTYGNACRQPDGMQVYN 159

RESULT 6
A25972
17k antigen precursor - Rickettsia rickettsii
C:Species: Rickettsia rickettsii
C:Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 20-Aug-1999
R:Anderson, B.E.; Regnery, R.L.; Carlone, G.M.; Tzianabos, T.; McNamee, J.E.; Fu, Z.Y.
J. Bacteriol. 169, 2385-2390, 1987
A:Title: Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia ricketts
A:Reference number: A25972; MUID:87222152
A:Accession: A25972

```

Db      121 STRATVQGVDSQIFR 136      :::| ||: |
RESULT 10
523787
outer membrane lipoprotein precursor - Yersinia enterocolitica
C:Species: Yersinia enterocolitica
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 29-Sep-1999
C:Accession: S23787
R:Baeumler, A. J.; Hanke, K.
J. Bacteriol. 174, 1029-1035, 1992
A:Title: A lipoprotein of Yersinia enterocolitica facilitates ferrioxamine uptake in
A:Reference number: S23786; MUID:92121089
A:Accession: S23787
A:Status: preliminary
A:Molecule type: DNA

```

A:Residues: 1-155 <BAE>
 A:Cross-references: EMBL:X50448; NID:948577; PIDN:CAA4297.1; PID:948579
 C:Superfamily: PAL cross-reacting lipoprotein

Query Match

Best Local Similarity 13.6%; Score 113.5; DB 2; Length 155;
 Matches 37; Conservative 19; Mismatches 43; Indels 51; Gaps 4;

QY 10 SLIIISVFLVGCAGN-----FSHQE-----VGAAT 34
 Db 7 AATAATTTTGCAANNNTTSGDFSAQKOVQTVYGTLLSYRPTTIGGDDNNVMGATG 66
 QY 35 GAVVGVAGQLFGKSGSRVMAIGAVLIGSKTIGOSMDQOKIKL----- 82
 Db 67 GAVLGGFLGNTVGGGTGSLATTAAGAVAGMAGCGVAGMANNITGVLEVRKDDGTTILV 126
 QY 83 --NLSLEKVKAGQVTRNRNDTGSYSVEP 110
 Db 127 VOKGQPFREFVGO--RVMLASGSTVTWSP 154

RESULT 11

882837

Conserved hypothetical protein XF0178 [Imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: B82837
 R:anonymous
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; PMID:20365717
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: B82837
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-257 <SIM>
 A:Cross-references: GB:AE003872; GB:AE003849; NID:99104975; PIDN:AAFR2991.1; GSPDB:GN001
 R:Simpson, A.J.G.; Reimach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 as-Neto, E.; Docena, M.R.P.; Camargo, L.E.A.; Carraro, D.M.; Carter, H
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitzajima, J.P.; Krieger, S.C.; Franco, M.C.; Frohm
 A:Authors: Martins, E.M.B.N.; Madalira, J.M.F.; Marino, C.L.; Marques, M.V.; Martins, F
 Rodrigues, V.; Rosa, A.J. de M.; de Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.
 M.; Tsuchioka, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Contents: annotation
 C:Accession: A59328
 A:Gene: XF0178

Query Match
 Best Local Similarity 13.4%; Score 112; DB 2; Length 257;
 Matches 20; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 31 GAATGAVGVAGQLFGKSGSRVMAIGAVLIGLIGSKI 70
 Db 105 GTGALITGGLGVGNQFGNGRRALTRAGAVAGVAGFIGNEV 144

RESULT 12

AD2596

lipa protein [Imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
 C:Accession: AD2596
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
 erage, G.; Gilliet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl
 Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Tung, M.; Krespan, W.; Perry, M.; Gordon-
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AD2696
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-142 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AA141986.1; PID:91739358; GSPDB:GN00186
 C:Experimental source: Strain C58 (Dupont)
 A:Gene: lipa
 A:Map position: circular chromosome

Query Match

Best Local Similarity 13.3%; Score 111.5; DB 2; Length 142;
 Matches 41; Conservative 21; Mismatches 60; Indels 29; Gaps 9;

QY 10 SLIIISVFLVGCAGNFSQEGATGAVGVAGQLFGKSGSRVMAIGAVLIGLIGSK 69
 Db 12 SLICVSM-LSAC-----TTGTTPAG--GSLFGR-SAGSTPTFLANLGGIGV-K 56
 QY 70 IGOSMDQOKIKL-----NLSLEKVKAGQVTRNRNDT-GNSYSVEPRTYORKNQRQ 124
 Db 57 SGVELDRGDQTKALEVEYKALETAPVGTPIVITGDDVKGGVAMAP---YQGN----- 107
 QY 125 QYCEPFQOKMAMGOKOETGTACROPDGR 155
 Db 108 QNCRQYSHLTIVDGRDTRVRGACRNDGSM 138

RESULT 13

D97478

lipa protein [Imported] - Agrobacterium tumefaciens (strain C58, Cereon)
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
 A:Reference number: A87359; PMID:11743194
 A:Accession: D97478
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-125 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK6781.1; PID:915155981; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C_1782
 A:Map position: circular chromosome

Query Match

Best Local Similarity 13.3%; Score 111; DB 2; Length 125;
 Matches 36; Conservative 18; Mismatches 61; Indels 16; Gaps 6;

QY 30 VQAATGAVGVAGQLFGKSGSRVMAIGAVLIGLIGSKTIGOSMDQOKIKL---NOS 85
 Db 2 LSACTTGTTRAGOSLFR-SAPSTPLANLGGIGV-KSGVELDRGDQTKALAEYKA 59
 QY 86 LEKVKAGQVTRNRNDT-GNSYSVEPRTYORKNQRQYCEPFQOKMAMGOKOET 144
 Db 60 LETAPVGTPIVITGDDVKGGVAMAP---YQGN-----QNCROYSHLTIVDGRDTRVR 110
 QY 145 GTACROPDGR 155
 Db 111 GAACRNDGSM 121

RESULT 14

S58234

lipa protein - Rhizobium leguminosarum
 C:Species: Rhizobium leguminosarum

C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 08-Oct-1999
C:Accession: S58234
R:Yeoman, K.H.; Delgado, M.J.; Downie, J.A.; Johnston, A.W.B.
submitted to the EMBL Data Library, July 1995
A:Reference number: S58232
A:Accession: S58234
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-139 <YEO>
A:Cross-references: EMBL:X89983; NID:g929821; PIDN:CA662010.1; PID:g929824

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:56:01 ; Search time 8.39378 Seconds

471.414 Million cell updates/sec

Title: US-09-677-374-2

Sequence: 1 MNRGCLGGSSLIISVFLVG.....IYGTACRQPDGRWQVISTEK 162

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters

231628

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 08

Listing first 45 summaries

Database : Issued_Patents_AA:*

```

1: /cgn2_6/ptodata/1/1aa/5a_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5b_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6a_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6b_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	81.5	9.7	651	4	US-08-556-978B-19	Sequence 19, App1
2	81.5	9.7	651	4	US-09-247-806-1	Sequence 1, App1
3	81.5	9.7	718	1	US-08-425-069-2	Sequence 2, App1
4	81.5	9.7	718	2	US-08-317-844B-2	Sequence 2, App1
5	81.5	9.7	747	3	US-09-034-177-3	Sequence 3, App1
6	78	9.3	865	1	US-07-803-633A-13	Sequence 13, App1
7	77.5	9.3	255	4	US-09-553-998-8	Sequence 8, App1
8	77	9.2	551	2	US-09-067-351-2	Sequence 2, App1
9	77	9.2	551	4	US-09-360-490-2	Sequence 2, App1
10	76.5	9.2	2516	3	US-08-374-077C-2	Sequence 2, App1
11	76.5	9.2	2516	4	US-08-895-590-2	Sequence 2, App1
12	76	9.1	1021	1	US-07-910-760-12	Sequence 12, App1
13	76	9.1	1021	1	US-08-440-519-12	Sequence 12, App1
14	76	9.1	1021	4	US-08-440-549-12	Sequence 12, App1
15	75	9.0	236	2	US-08-190-199A-65	Sequence 65, App1
16	75	9.0	240	2	US-08-956-047-25	Sequence 25, App1
17	74.5	8.9	913	1	US-08-220-151-6	Sequence 6, App1
18	74.5	8.9	913	3	US-08-413-118-6	Sequence 6, App1
19	74.5	8.9	913	3	US-08-473-446-6	Sequence 6, App1
20	74.5	8.9	913	4	US-09-232-658A-2	Sequence 2, App1
21	74	8.9	970	1	US-08-375-709-7	Sequence 7, App1
22	74	8.9	970	1	US-08-752-929-7	Sequence 7, App1
23	74	8.9	970	4	US-09-090-793-5	Sequence 5, App1
24	73.5	8.8	546	2	US-09-067-351-1	Sequence 1, App1
25	73.5	8.8	546	4	US-09-360-490-1	Sequence 1, App1
26	73.5	8.8	733	4	US-08-464-700-2	Sequence 2, App1
27	73	8.7	142	1	US-08-556-823-2	Sequence 2, App1

28	73	8.7	25.4	4	US-09-128-450-26	Sequence 26, Appl
29	73	8.7	25.4	4	US-09-823-494-26	Sequence 26, Appl
30	73	8.7	60.6	4	US-09-247-806-6	Sequence 6, Appl
31	72.5	8.7	47.9	4	US-09-177-349-3	Sequence 3, Appl
32	72.5	8.7	54.7	1	US-08-340-203A-3	Sequence 3, Appl
33	72.5	8.7	54.7	2	US-08-452-567-3	Sequence 3, Appl
34	72.5	8.7	54.7	2	US-08-452-427-3	Sequence 3, Appl
35	72.5	8.7	54.7	3	US-09-085-407-3	Sequence 3, Appl
36	72.5	8.7	73.1	2	US-08-911-364-1	Sequence 1, Appl
37	72.5	8.7	73.2	2	US-08-678-039A-0	Sequence 10, Appl
38	72	8.6	10.1	4	US-08-556-978B-20	Sequence 20, Appl
39	72	8.6	10.1	4	US-08-556-978B-22	Sequence 22, Appl
40	72	8.6	10.1	4	US-08-556-978B-82	Sequence 62, Appl
41	72	8.6	10.1	4	US-09-247-806-3	Sequence 3, Appl
42	72	8.6	10.1	4	US-09-247-806-5	Sequence 5, Appl
43	72	8.6	10.1	4	US-09-247-806-7	Sequence 7, Appl
44	72	8.6	60.4	4	US-08-556-978B-53	Sequence 63, Appl
45	72	8.6	60.4	4	US-08-556-978B-51	Sequence 21, Appl

ALIGNMENTS

RESULT 1

; Sequence 19, Application US/08556978B

; GENERAL INFORMATION:

TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED

NUMBER OF SEQUENCES: 107

ADDRESSEE: E. I. DU PONT

CITY: WILMINGTON

COUNTRY: UNITED STATES OF AMERICA

COMPUTER READABLE FORM

COMPUTER: IBM PC COMPATIBLE

SOFTWARE: MICROSOFT WORD FOR WINDOWS 95

APPLICATION NUMBER: US/08/556,978B

CLASSIFICATION: 435

APPLICATION NUMBER: 08/077,600

ATTORNEY/AGENT INFORMATION

REGISTRATION NUMBER: 33,692

TELECOMMUNICATION INFORMATION:

TELEFAX: 302-773-0164

SEQUENCE CHARACTERISTICS:

TYPE: amino acid ;

TOPOLOGY: unknown

US-08-556-978B-19

Query Match	9.78;	Score 81.5;	DB 4;	Length 651;

Matches 25; Conservative 3; Mismatches 17; Indels 23; Gaps

QY 28 QEVGATGAVG---GVAGQ-----LFGKSGRVAMIGGAVLGG 64

Db 491 OGAGAAAAVAGAGGEGIRGAGAGGCGYGLGSGSGRGGLGCGAGAGAAAAAGAGAGGG 550
QY 65 LIGSKTGO 72
Db 551 LGGGAGG 558

RESULT 2
US-09-247-806-1
Sequence 1, Application US/09247806
Patent No. 6280747
GENERAL INFORMATION:
APPLICANT: PHILLIPE, Michel
APPLICANT: GARSON, Jean-Claude
APPLICANT: ARAUDEAU, Jean-Pierre
TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
FILE REFERENCE: 6388-0365-0
CURRENT APPLICATION NUMBER: US/09/247,806
EARLIER FILING DATE: 1999-02-11
EARLIER FILING DATE: 1998-02-11
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 1
LENGTH: 651
TYPE: PRT
ORGANISM: Nephila clavipes
US-09-247-806-1

Query Match
Best Local Similarity 9.7%; Score 81.5; DB 4; Length 651;
Matches 25; Conservative 3; Mismatches 17; Indels 23; Gaps 2;

QY 28 QEWGATGAVG---GVAGQ-----LFGKSGRVAMATGAVLGG 64
Db 491 OGAGAAAAVAGAGGEGIRGAGAGGCGYGLGSGSGRGGLGCGAGAGAAAAAGAGAGGG 550
QY 65 LIGSKTGO 72
Db 551 LGGGAGG 558

RESULT 3
US-08-425-069-2
Sequence 2, Application US/08425069
Patent No. 5728810
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5728810th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,069
FILING DATE: 19-APR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-069-2

Query Match
Best Local Similarity 9.7%; Score 81.5; DB 1; Length 718;
Matches 25; Conservative 3; Mismatches 17; Indels 23; Gaps 2;

QY 28 QEWGATGAVG---GVAGQ-----LFGKSGRVAMATGAVLGG 64
Db 491 OGAGAAAAVAGAGGEGIRGAGAGGCGYGLGSGSGRGGLGCGAGAGAAAAAGAGAGGG 550
QY 65 LIGSKTGO 72
Db 551 LGGGAGG 558

RESULT 4
US-08-317-844B-2
Sequence 2, Application US/08317844B
Patent No. 5989894
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5989894th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,844B
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 241-1300
TELEFAX: (703) 241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-317-844B-2

Query Match 9.7%; Score 81.5; DB 2; Length 718;

Best Local Similarity 36.8%; Pred. No. 0.91;
Matches 25; Conservative 3; Mismatches 17; Indels 23; Gaps 2;

OY 28 QEWGATGAVG---GVAGQ-----LFGKSGRVAMATGAVLGG 64

Db 491 OGAGAAAAAAGAGGEGIRGAGGAGGGYGGILGSGSGRGGILGGGAGAAAAAAGGAGGCG 550

OY 65 LIGSKIGQ 72

Db 551 LGGGAGQ 558

RESULT 5

US-09-034-177-3
Sequence 3, Application US/09034177
Patent No. 6127146

GENERAL INFORMATION:

APPLICANT: Lal, Preeti

APPLICANT: Guegler, Karl J.

APPLICANT: Corley, Neil C.

TITLE OF INVENTION: HUMAN FIBROUS PROTEIN

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSO for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/034,177

FILING DATE: HEREMITH

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0486 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

TELEX:

INFORMATION FOR SEQ. ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 747 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: GI 1174414

US-09-034-177-3

Query Match 9.7%; Score 81.5; DB 3; Length 747;

Best Local Similarity 36.8%; Pred. No. 0.96;

Matches 25; Conservative 3; Mismatches 17; Indels 23; Gaps 2;

OY 28 QEWGATGAVG---GVAGQ-----LFGKSGRVAMATGAVLGG 64

Db 491 OGAGAAAAAAGAGGEGIRGAGGAGGGYGGILGSGSGRGGILGGGAGAAAAAAGGAGGCG 550

OY 65 LIGSKIGQ 72

Db 551 LGGGAGQ 558

RESULT 6

US-07-803-633A-13
Sequence 13, Application US/07803633A
Patent No. 5365025

GENERAL INFORMATION:

APPLICANT: NAZERIAN, Keyvan

APPLICANT: LEE, Lucy F.

APPLICANT: YANAGIDA, No. 5369025oru

APPLICANT: OGAMA, Ryohel

APPLICANT: LI, Yi

TITLE OF INVENTION: RECOMBINANT FOMLPX VACCINE FOR

PROTECTION AGAINST MAREK'S DISEASE

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH

STREET: 301 No. 5369025th Washington Street

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentia Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/803,633A

FILING DATE: 19911210

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1644-103P

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 241-1300

TELEFAX: (703) 241-2848

TELEX: 248345

INFORMATION FOR SEQ. ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 865 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-803-633A-13

Query Match 9.3%; Score 78; DB 1; Length 865;

Best Local Similarity 25.4%; Pred. No. 2.9;

Matches 36; Conservative 30; Mismatches 48; Indels 28; Gaps 7;

OY 17 FIVGCAQNFPSRQ-EVGAATGAVVAGVAGQLEFGKSGRV-----AMATGAVLGLIG 67

Db 688 FMNGLAELFNGMGVQVGAIRKVVAGAGAVTSTISVSAPFMSNPFGLALGLIAGLVA 747

OY 68 SKIGOSMDODDKIKLN-----QSLKVKAGVYTRMRNPDGNSVSEVPRTYQRYNKQ 120

Db 748 AFL--AIFYVVKLSNPMKALYPTTEVLKA-QATRELHEESD-----LERSID 796

OY 121 ERROQYCRE-FQOKAMTAGQKQ 141

Db 797 ERKLEAREMIKYWALVSAER 818

RESULT 7

US-09-533-498-8
Sequence 8, Application US/09553498
Patent No. 6309861

GENERAL INFORMATION:

APPLICANT: Ambrosius, Dorothee

APPLICANT: Rudolph, Rainer

APPLICANT: Schaeffner, Joerg

APPLICANT: Schwartz, Elisabeth

TITLE OF INVENTION: Process for the production of naturally folded and secreted pr

FILE REFERENCE: Case 20379

Tue Oct 29 08:30:56 2002

us-09-677-374-2.ra1

CURRENT APPLICATION NUMBER: US/09/553,498
CURRENT FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: EP99107412.1
NUMBER OF SEQ ID NOS: 10
LENGTH: 255
TYPE: PRP
ORGANISM: E. coli
US-09-553-498-8

Query Match
Best Local Similarity 20.9%; Score 77.5; DB 4; Length 255;
Matches 37; Conservative 26; Mismatches 81; Indels 31; Gaps 7;

QY 6 LOGSLITISVFLVGCANFSGEVAATGAVV---GVAQGLFGKSGRVAMATGAVLIG 61
DB 83 LOMTSLREPTAMVYCARDYGAT-MGQGTTVVSSGGSGGSGGSDIELTSPAI 141
QY 62 LGLLISKIQGSMDOODKIK-LN-----QSLKRVAGQVTRMRNPDTCNSY 106
DB 142 MSASPGKRVMTCSASSRVYNNMFQOKSGTSPKRWIYDTSKLSGVPARFSGSGGTSY 201
QY 107 SVEPRTYORYNKOEPRQOYCEFOQKMI--AGQKEIYGTACRQPDGRMQVISTE 161
DB 202 SL-----TISMEADPAATYTCQWSSNPLTFGAGTKLEKRAAADQ-----KLISEE 249

RESULT 8
US-09-067-351-2
Sequence 2, Application US/09067351
Patent No. 5994081
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN KERATINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/067,351
FILING DATE: Herewith
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
TELECOMMUNICATION INFORMATION: PF-0511 US
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 551 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: KERANOT02
CLONE: 2029060
US-09-067-351-2

Query Match
Best Local Similarity 22.4%; Score 77; DB 2; Length 551;
Matches 28; Conservative 28; Mismatches 53; Indels 16; Gaps 2;

QY 8 GSSLITISVFLVGCANFSGEVAATGAVVAGVAGLFGKSGRVAMATGAVLIGLIG 67
DB 54 GASFGSRSLYNLGAKRVSLNGCGSSCRSGFGGRASNGFVNSG---FGYGGVGGGFGSG 110
QY 68 SK-----TGSMDOODKIKINOSLEKRVAGQVTRMRNPDTCNSYSEVPRTY 114
DB 111 PSPFVCPGQIDQEVYVNSLPLPHLIDIPITQVAREEIOIKTLNKFSTIDKVFEL 170
QY 115 QQRNK 119
DB 171 EQQNK 175

RESULT 9
US-09-360-490-2
Sequence 2, Application US/09360490
Patent No. 6221843
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN KERATINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/360,490
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/067,351
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
TELECOMMUNICATION INFORMATION: PF-0511 US
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 551 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: KERANOT02
CLONE: 2029060
US-09-360-490-2

Query Match
Best Local Similarity 9.2%; Score 77; DB 4; Length 551;
Matches 28; Conservative 28; Mismatches 53; Indels 16; Gaps 2;

QY 8 GSSLITISVFLVGCANFSGEVAATGAVVAGVAGLFGKSGRVAMATGAVLIGLIG 67
DB 54 GASFGSRSLYNLGAKRVSLNGCGSSCRSGFGGRASNGFVNSG---FGYGGVGGGFGSG 110

Qy 68 SK-----IGSMDOODKIKLNOSLEKVKAGVTRMRNPDTCNSVEPVRTY 114
Db 111 PSPVPCPGIOEVYNOSLLTPLHLQIDPTIGRVAREERQIKTNKFTSIDKVRFL 170
Qy 115 QRYNK 119
Db 171 EQQNK 175

RESULT 10
US-08-374-077C-2
Sequence 2, Application US/08374077C
Patent No. 6027912
GENERAL INFORMATION:
APPLICANT: Hall, Linda M.
APPLICANT: Ren, Dejian
APPLICANT: Zheng, Wei
APPLICANT: Dubald, Manuel Marcel Paul
TITLE OF INVENTION: Genes Encoding an Invertebrate Alpha1
TITLE OF INVENTION: Calcium Channel Subunit
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/374,077C
FILING DATE: 19-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm M.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 022650-264
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2516 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-374-077C-2

Query Match 9.2%; Score 76.5; DB 3; Length 2516;
Best Local Similarity 24.8%; Pred. No. 17;
Matches 32; Conservative 18; Mismatches 66; Indels 13; Gaps 4;

Qy 30 VCAATGAVVGAGQLFGKSGRYVMAIGG--AVLGGLGSKIGQS---MDQODKIKLN 83
Db 2389 IGSSNGSIFGSGAGLGGAGSGVG--GLGSSSIRNAFGSGSGPSLSPOHQPYSGTLN 2447
Qy 84 -----QSELEKVKAGVTRMRNPDTCNSYSVEFVRYQRYNKQERRQOYCREFQOKAMIA 137
Db 2448 SPPIPNRLRRVATVTNNKRSQVSONSSSLNVRANANSQMNNSPTGQPVQOOSPLR 2507
Qy 138 GOKOEIYGT 146
Db 2508 GQGNQTYSS 2516

RESULT 11
US-08-895-590-2
Sequence 2, Application US/08895590
Patent No. 6207410

GENERAL INFORMATION:
APPLICANT: Hall, Linda M.
APPLICANT: Ren, Dejian
APPLICANT: Zheng, Wei
APPLICANT: Dubald, Manuel Marcel Paul
TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895,590
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/374,888
FILING DATE: 19-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm M.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 022650-263
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2516 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-895-590-2

Query Match 9.2%; Score 76.5; DB 4; Length 2516;
Best Local Similarity 24.8%; Pred. No. 17;
Matches 32; Conservative 18; Mismatches 66; Indels 13; Gaps 4;

Qy 30 VCAATGAVVGAGQLFGKSGRYVMAIGG--AVLGGLGSKIGQS---MDQODKIKLN 83
Db 2389 IGSSNGSIFGSGAGLGGAGSGVG--GLGSSSIRNAFGSGSGPSLSPOHQPYSGTLN 2447
Qy 84 -----QSELEKVKAGVTRMRNPDTCNSYSVEFVRYQRYNKQERRQOYCREFQOKAMIA 137
Db 2448 SPPIPNRLRRVATVTNNKRSQVSONSSSLNVRANANSQMNNSPTGQPVQOOSPLR 2507
Qy 138 GOKOEIYGT 146
Db 2508 GQGNQTYSS 2516

RESULT 12
US-07-910-760-12
Sequence 12, Application US/07910760
Patent No. 5683864
GENERAL INFORMATION:
APPLICANT: Houghton, Michael
APPLICANT: Kuo, Qui-Lim
APPLICANT: Kuo, George
TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: P.O. Box 8097 (Int. Prop. R-440)
CITY: Emeryville

STATE: CA
 COUNTRY: U.S.A.
 ZIP: 94662-8097
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/910,760
 FILING DATE: 07-JUL-1992
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Blackburn Esq., Robert P.
 REGISTRATION NUMBER: 30,447
 REFERENCE/DOCKET NUMBER: 0101.002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (510) 601-2702
 TELEFAX: (510) 655-3542
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1021 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-910-760-12

Query Match
 Best Local Similarity 9.1%; Score 76; DB 1; Length 1021;
 Matches 42; Conservative 24; Mismatches 63; Indels 72; Gaps 9;

QY 31 GAATGAVGVAGAGQLFGK-GSGRVAMAI-----GGAVALGLIGSKI--GQSMDOQDKIKLN 83
 DB 792 GAATAFVAGAGLAGAAGISVGLKVLIDLAGYAGVAGALVAFKIMSGVPTEDLVNLL 851
 QY 84 QSL-----EKVAKQ-VTRWRN-----PDGNSYSVEPVRYTORVKNQ 120
 DB 852 PALLSPALVGVGCAILRRHVGPBGAVQVMNRLIAFASRGNHVSPPNSSTNPKPKK 911
 QY 121 ERRO-----OYCFEPQOKAMITAGOKOEIYGTAC 148
 DB 912 NKRNTNRRODVKFFPGGQIVGVYLLPRRGPRLGVATRTKTSERSOPRGRPOPI--PKA 969
 QY 149 ROPDGR-----WOYISTE 161
 DB 970 RRPGRRTWAPGYPWPLYGNE 990

RESULT 13
 US-08-440-519-12
 Sequence 12, Application US/08440519
 Patent No. 5712087
 GENERAL INFORMATION:
 APPLICANT: Houghton, Michael
 APPLICANT: Choo, Qui-Lim
 TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Chiron Corporation
 STREET: P.O. Box 8097 (Int. Prop. R-440)
 CITY: Emeryville
 STATE: CA
 COUNTRY: U.S.A.
 ZIP: 94662-8097
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/440,519

FILING DATE: 12-MAY-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/910,760
 FILING DATE: 07-JUL-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Blackburn Esq., Robert P.
 REGISTRATION NUMBER: 30,447
 REFERENCE/DOCKET NUMBER: 0101.002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (510) 601-2702
 TELEFAX: (510) 655-3542
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1021 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-440-519-12

Query Match
 Best Local Similarity 9.1%; Score 76; DB 1; Length 1021;
 Matches 42; Conservative 24; Mismatches 63; Indels 72; Gaps 9;

QY 31 GAATGAVGVAGAGQLFGK-GSGRVAMAI-----GGAVALGLIGSKI--GQSMDOQDKIKLN 83
 DB 792 GAATAFVAGAGLAGAAGISVGLKVLIDLAGYAGVAGALVAFKIMSGVPTEDLVNLL 851
 QY 84 QSL-----EKVAKQ-VTRWRN-----PDGNSYSVEPVRYTORVKNQ 120
 DB 852 PALLSPALVGVGCAILRRHVGPBGAVQVMNRLIAFASRGNHVSPPNSSTNPKPKK 911
 QY 121 ERRO-----OYCFEPQOKAMITAGOKOEIYGTAC 148
 DB 912 NKRNTNRRODVKFFPGGQIVGVYLLPRRGPRLGVATRTKTSERSOPRGRPOPI--PKA 969
 QY 149 ROPDGR-----WOYISTE 161
 DB 970 RRPGRRTWAPGYPWPLYGNE 990

RESULT 14
 US-08-440-549-12
 Sequence 12, Application US/08440549
 Patent No. 6312889
 GENERAL INFORMATION:
 APPLICANT: Houghton, Michael
 APPLICANT: Choo, Qui-Lim
 TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Chiron Corporation
 STREET: P.O. Box 8097 (Int. Prop. R-440)
 CITY: Emeryville
 STATE: CA
 COUNTRY: U.S.A.
 ZIP: 94662-8097
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/440,549
 FILING DATE: 12-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/910,760
 FILING DATE: 07-JUL-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Blackburn Esq., Robert P.


```
;
;   REGISTRATION NUMBER: 30,447
;   REFERENCE/DOCKET NUMBER: 0101.002
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (510) 601-2702
;   TELEFAX: (510) 655-3542
;   INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1021 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;
US-08-440-549-12

Query Match
Best local Similarity 20.9%; Pred. No. 6:
Matches 42; Conservative 24; Mismatches 63; Indels 72; Gaps 9;

OY 31 GAATGAVGVGAGQLFGK-GSGRVAMAI-----GCAVLGLIGSKI--GQSMDOQDKIKLN 83
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 792 GAATATVAGAGLGAAGVGLGKVLIDILAGYAGAGVAGVAFKIMSGEVPSTEDLVNLL 851
OY 84 QSL-----EKVAKGQ-VTRRN-----PDTGNSYSVFPVRYQRYNKQ 120
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Db 852 PALLSPGALVGVYCAAILRRHVPGEGAVQMMNRLIAFASRCNNHVSFGNSSTNPKPOKK 911
OY 121 ERQ-----QYCREFOQKAMIAQKQKEIYTAC 148
    :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db 912 NKRNRNRPPQDVKPEGGQIVGVYLLPRRPRLVATRKTSERSOPRGRQPI--PKA 969
OY 149 RPPDGR-----MOVISTE 161
    |:|:|  |:|:|  |:|:|  |:|:|  |:|:|  |:|:|  |:|:|  |:|:|  |:|:|  |:|:|
Db 970 RREPRTMAQPCYPMPLYGNE 990

RESULT 15
US-08-190-199A-65
; Sequence 65, Application US/08190199A
; Patent No. 5830663
; GENERAL INFORMATION:
;   APPLICANT: EMBLETON, Michael J.
;   APPLICANT: GOROCHOV, Guy
;   APPLICANT: JONES, Peter T.
;   APPLICANT: WINTER, Gregory P.
;   TITLE OF INVENTION: TREATMENT OF CELL POPULATIONS
;   NUMBER OF SEQUENCES: 70
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: PILLSBURY MADISON & SUTRO, L.L.P.
;   STREET: 1100 New York Avenue, N.W.
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20005-3918
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Microsoft word
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/190,199A
;   FILING DATE: 13-JUL-1994
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/GB92/01483
;   FILING DATE: 10-AUG-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: GB 9212419.7
;   FILING DATE: 11-JUN-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: GB 9117352.6
;   FILING DATE: 10-AUG-1991
;   INFORMATION FOR SEQ ID NO: 65:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 236 amino acids
```

```
;
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;
US-08-190-199A-65

Query Match
Best local Similarity 19.6%; Pred. No. 1,1;
Matches 31; Conservative 27; Mismatches 76; Indels 24; Gaps 5;

OY 6 LOGSSLIITISVPLVGAQNFNR---QEVGATAVGVGAGQLFGKSGSRVAMAIAGAVL 62
    ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|
Db 81 LQMTSLRSEDDTAMYYCARDYGAYWGGTTLVTSAGGGSGGGGGSQIVLTGSPAIM 140
OY 63 GGLIGKIGQSMQDQKIK-LN-----QSLKVRKAGVTRRNPDGNSYS 107
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 141 SASPGEKVTWTCASSSVRYMNFQOKSGTSPKRWIYDTKSLSSGVAPRFGSGSGSTYS 200
OY 108 VEPVRYQRYNKOERRQYCREFOQKAMI--AQOKOI 143
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 201 L----TSSMEADATYYQQWSSNPLTFGAGTKIEL 234

Search completed: October 27, 2002, 11:00:27
Job time : 10.3938 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:55:57 : Search time 21.8238 Seconds
(without alignments)
824.509 Million cell updates/sec

Title: US-09-677-374-2

Perfect score: 836

Sequence: 1 MNRGCLGSGSLIISVPLVG.....IYGFACRQPDGRWQVISTEK 162

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_032802.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	836	100.0	162	AA1980	Piscirickettsia sa
2	836	100.0	162	AA1981	OspA antigen amino
3	815	97.5	161	AA1982	Optimised OspA pro
4	815	97.5	256	AA1983	CITE2 OspA constru
5	112	13.4	20	AA1984	OspA B-cell epitop
6	108	12.9	224	AA1985	Moraxella catarin
7	105.5	12.6	223	AA1986	Porphyromonas ging
8	105.5	12.6	220	AA1987	Porphyromonas ging
9	102	12.2	154	AA1988	PBOBP-2 gene prod.
10	98	11.7	309	AA1989	Novel human diagno
11	91	10.9	528	AA1990	Spider recombinant

12	86.5	10.3	2309	22	ABB66232
13	82	9.8	666	22	ABB58019
14	81.5	9.7	651	20	AAV40097
15	81.5	9.7	718	12	AA14308
16	81.5	9.7	718	19	AAW53346
17	81.5	9.7	718	21	AAV59070
18	80.5	9.6	102	22	AAW41943
19	80.5	9.6	302	22	AAW40157
20	80.5	9.6	542	22	ABB65790
21	80.5	9.6	542	22	ABB65791
22	80.5	9.6	542	22	ABB70501
23	80	9.6	116	19	AAV11028
24	80	9.6	2017	22	ABG06301
25	80	9.6	2599	21	AAV75098
26	79	9.4	618	21	AA556803
27	78	9.3	646	18	AAW27178
28	78	9.3	865	14	AA130169
29	77.5	9.3	255	21	AA11398
30	77.5	9.3	255	22	AA174199
31	77.5	9.3	255	22	AA170769
32	77.5	9.3	255	22	AA172020
33	77.5	9.3	900	22	ABG28648
34	77	9.2	147	22	ABG9847
35	77	9.2	429	10	AA190064
36	77	9.2	551	21	AA152398
37	77	9.2	1251	22	ABB61254
38	76.5	9.2	514	22	AAU36520
39	76.5	9.2	2516	17	AAW01875
40	76.5	9.2	2516	17	AAW01884
41	76.5	9.2	2516	22	ABB61076
42	76	9.1	300	11	AA105878
43	76	9.1	324	22	AA176751
44	76	9.1	346	22	AA192057
45	76	9.1	346	22	AA176750

ALIGNMENTS

RESULT 1	
AA198025	
ID	AA198025 standard; Protein; 162 AA.
XX	
AC	AA198025.
XX	
DT	15-JAN-2002 (first entry)
XX	
DE	Piscirickettsia salmonis polypeptide p10.6.
XX	
KW	Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
KW	septicaemia; SRS; surface antigen; vaccine; antibacterial; fish;
KW	ATCC VR-1361.
XX	
OS	Piscirickettsia salmonis.
XX	
PN	W0200168865-A2.
XX	
PD	20-SEP-2001.
XX	
PF	12-MAR-2001; 2001MO-GB01055.
XX	
PR	11-MAR-2000; 2000GB-0005838.
PR	01-JUL-2000; 2000GB-0016080.
PR	01-JUL-2000; 2000GB-0016082.
PR	29-JUL-2000; 2000GB-0018599.
PA	(AQUA-) AQUA HEALTH EURO LTD.
XX	
PI	Simard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;
PI	Burzio L;
XX	
DR	WPI: 2001-639050/73.
DR	N-PDB; AA179040.

XX New nucleic acids encoding an amino acid sequence homologous to the
PT surface antigen present on piscirickettsia salmonis are useful to
PT protect fish against piscirickettsiosis

Claim 6; Fig 5; 25pp; English.

The invention relates to nucleic acid sequences and the encoded protein
CC of a least part of the surface antigen present on piscirickettsia
CC salmonis for production of a vaccine with antibacterial activity to
CC protect fish against P. salmonis which causes piscirickettsiosis, also
CC known as salmonid rickettsial septicaemia.

Sequence 162 AA:

Query Match 100.0%; Score 836; DB 22; Length 162;
Best Local Similarity 100.0%; Pred. No. 3.2e-80;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGCLQGSSLLIITSVLVCAQNFSSROEVGAATGAVGVAGOLFSGSGRVAMAIGCA 60
DB 1 MNRGCLQGSSLLIITSVLVCAQNFSSROEVGAATGAVGVAGOLFSGSGRVAMAIGCA 60
QY 61 VLGLGSLKIGSGMDQDKIKLNSLEKVKAGQVTRMRNPDTGNSYSVEPVRTYQRYNKQ 120
DB 61 VLGLGSLKIGSGMDQDKIKLNSLEKVKAGQVTRMRNPDTGNSYSVEPVRTYQRYNKQ 120
QY 121 ERROQYCREFOQKAMTAGOKOEIYGTACROPDGRMQVISTEK 162
DB 121 ERROQYCREFOQKAMTAGOKOEIYGTACROPDGRMQVISTEK 162

RESULT 2

ID AAB81126 standard; Protein; 162 AA.
AC AAB81126;
XX

DT 11-JUL-2001 (first entry)
XX

DE Ospa antigen amino acid sequence.
XX

KW Poikilothermic fish; piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
XX SRS.

XX Piscirickettsia salmonis.
OS

Key Location/Qualifiers
FT 110..129
FT Region /label=B_cell_epitope
XX

CA2281913-A1.
XX

17-MAR-2001.
XX

PF 17-SEP-1999; 99CA-2281913.
XX

PR 17-SEP-1999; 99CA-2281913.
XX

PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZV/) KUZV M A.
XX

PI Kay WW, Burian J, Kuzyk MA;
XX

DR WPI; 2001-316844/34.
XX

PT N-PSDB; AAF86246.
XX
PT Method for protecting poikilothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
XX vaccine containing the Ospa protein of piscirickettsia salmonis

PS Example 2; Fig 2B; 35pp; English.
XX

CC This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly poikilothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents P. salmonis Ospa protein. An Ospa protein with an
CC antibody response.

Sequence 162 AA:

Query Match 100.0%; Score 836; DB 22; Length 162;
Best Local Similarity 100.0%; Pred. No. 3.2e-80;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGCLQGSSLLIITSVLVCAQNFSSROEVGAATGAVGVAGOLFSGSGRVAMAIGCA 60
DB 1 MNRGCLQGSSLLIITSVLVCAQNFSSROEVGAATGAVGVAGOLFSGSGRVAMAIGCA 60
QY 61 VLGLGSLKIGSGMDQDKIKLNSLEKVKAGQVTRMRNPDTGNSYSVEPVRTYQRYNKQ 120
DB 61 VLGLGSLKIGSGMDQDKIKLNSLEKVKAGQVTRMRNPDTGNSYSVEPVRTYQRYNKQ 120
QY 121 ERROQYCREFOQKAMTAGOKOEIYGTACROPDGRMQVISTEK 162
DB 121 ERROQYCREFOQKAMTAGOKOEIYGTACROPDGRMQVISTEK 162

RESULT 3

ID AAB81127 standard; Protein; 161 AA.
AC AAB81127;
XX

DT 11-JUL-2001 (first entry)
XX

DE Optimised Ospa protein 17E2 amino acid sequence.
XX

KW Poikilothermic fish; piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
XX SRS.

XX Piscirickettsia salmonis.
OS Synthetic.

Key Location/Qualifiers
FT 109..128
FT Region /label=B_cell_epitope
XX

CA2281913-A1.
XX

PD 17-MAR-2001.
XX

PF 17-SEP-1999; 99CA-2281913.
XX

PR 17-SEP-1999; 99CA-2281913.
XX

PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZV/) KUZV M A.
XX

PI Kay WW, Burian J, Kuzyk MA;
XX

DR WPI; 2001-316844/34.
XX

PT N-PSDB; AAF86247.
XX
PT Method for protecting poikilothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a

PT vaccine containing the Ospa protein of *Piscirickettsia salmonis* -
XX
XX Disclosure: Fig 5; 35pp; English.
CC This invention relates to a method for the protection against infection
CC of a polkillothermic fish by the bacterial pathogen, *Piscirickettsia*
CC *salmonis*. The method comprises administering an immunogenic amount of a
CC *P. salmonis* specific antigen termed Ospa, or an immunogenic fragment of
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkillothermic fish, against the bacterial pathogen
CC *P. salmonis*. The method is also useful for protecting against salmonid
CC rickettsial septicæmia (SRS) and other rickettsial diseases. The present
CC sequence represents optimised *P. salmonis* Ospa protein 17E2. The DNA
CC encoding Ospa 17E2 (AA86247) has been optimised for expression in
CC *Escherichia coli*. An Ospa protein with an N-terminal fusion partner is
CC used in a vaccine to create an anti-Ospa antibody response.
XX
SQ Sequence 161 AA:
Query Match 97.5%; Score 815; DB 22; Length 161;
Best Local Similarity 98.8%; Pred. No. 5, 1e-78;
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 3 RCGLOGSSLLIIISVFLVGCANFSROEVGAATGAVVGAVAGOLFSGKSGRVAMAGAVL 62
DB 2 RCGLOGSSLLIIISVFLVGCANFSROEVGAATGAVVGAVAGOLFSGKSGRVAMAGAVL 61
OY 63 GGLIGSKIGSMQODKIKLNOSLEKVKAGQVTRWRNPDTGNSYSVEPRTYORYNKQER 122
DB 62 GGLIGSKIGSMQODKIKLNOSLEKVKAGQVTRWRNPDTGNSYSVEPRTYORYNKQER 121
OY 123 ROQYCREFOOKAMIAAGOKETIGTACROPDGRMOWISTEK 162
DB 122 ROQYCREFOOKAMIAAGOKETIGTACROPDGRMOWISTEK 161
RESULT 4
AAB81128
ID AAB81128 standard; Protein; 256 AA.
XX
AC AAB81128;
XX
DT 11-JUL-2001 (first entry)
XX
DE C17E2 Ospa construct with N-terminal fusion partner.
XX
KW Polkillothermic fish; *Piscirickettsia salmonis*; rickettsial pathogen;
KW vaccine; Ospa; salmonid rickettsial septicæmia; rickettsial disease;
KW SRS; 17E2; fusion construct.
XX
OS *Piscirickettsia salmonis*.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Region 1..95 /label= Undefined_N-terminal_fusion_partner
FT 96..256 /label= C17E2_Ospa
FT /note= "Product of Ospa gene optimised for expression in
FT *Escherichia coli*."
XX
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
XX (BURI/) BURIAN J.
XX (KUZV/) KUZYK M A.
XX

PI Kay WM, Burian J, Kuzyk MA;
XX
XX WPI: 2001-316844/34.
DR N-PDB; AAF86248.
XX
XX Method for protecting polkillothermic fish against salmonid rickettsial
PT septicæmia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of *Piscirickettsia salmonis* -
XX
XX Example 4; Fig 5; 35pp; English.
XX
PS This invention relates to a method for the protection against infection
CC of a polkillothermic fish by the bacterial pathogen, *Piscirickettsia*
CC *salmonis*. The method comprises administering an immunogenic amount of a
CC *P. salmonis* specific antigen termed Ospa, or an immunogenic fragment of
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkillothermic fish, against the bacterial pathogen
CC *P. salmonis*. The method is also useful for protecting against salmonid
CC rickettsial septicæmia (SRS) and other rickettsial diseases. The present
CC sequence represents the amino acid sequence of C17E2, a *P. salmonis* Ospa
CC construct optimised for expression in *Escherichia coli*, fused to an
CC undefined N-terminal fusion partner. The fusion protein is used in a
CC vaccine to create an anti-Ospa antibody response.
XX
SQ Sequence 256 AA:
Query Match 97.5%; Score 815; DB 22; Length 256;
Best Local Similarity 98.8%; Pred. No. 9, 3e-78;
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 3 RCGLOGSSLLIIISVFLVGCANFSROEVGAATGAVVGAVAGOLFSGKSGRVAMAGAVL 62
DB 97 RCGLOGSSLLIIISVFLVGCANFSROEVGAATGAVVGAVAGOLFSGKSGRVAMAGAVL 156
OY 63 GGLIGSKIGSMQODKIKLNOSLEKVKAGQVTRWRNPDTGNSYSVEPRTYORYNKQER 122
DB 157 GGLIGSKIGSMQODKIKLNOSLEKVKAGQVTRWRNPDTGNSYSVEPRTYORYNKQER 216
OY 123 ROQYCREFOOKAMIAAGOKETIGTACROPDGRMOWISTEK 162
DB 217 ROQYCREFOOKAMIAAGOKETIGTACROPDGRMOWISTEK 256
RESULT 5
AAB81130
ID AAB81130 standard; Peptide; 20 AA.
XX
AC AAB81130;
XX
DT 11-JUL-2001 (first entry)
XX
DE Ospa B-cell epitope peptide #2.
XX
KW Polkillothermic fish; *Piscirickettsia salmonis*; rickettsial pathogen;
KW vaccine; Ospa; salmonid rickettsial septicæmia; rickettsial disease;
KW SRS; antibody.
XX
OS *Piscirickettsia salmonis*.
OS
XX
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
XX (BURI/) BURIAN J.
XX (KUZV/) KUZYK M A.
XX
XX Kay WM, Burian J, Kuzyk MA;
XX

DR WPI: 2001-316844/34.

XX Method for protecting poikilothermic fish against salmonid rickettsial
PT septicemia and other rickettsial diseases comprises administering a
XX vaccine containing the Ospa protein of *Piscirickettsia salmonis* -
PS Example 2; Page 17; 35pp; English.

XX This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*
CC salmonis. The method comprises administering an immunogenic amount of a
CC *P. salmonis* specific antigen termed Ospa, or an immunogenic amount of a
CC Ospa in the form of a vaccine. The method is used for protecting fragment of a
CC particularly poikilothermic fish, against the bacterial pathogen
CC *P. salmonis*. The method is also useful for protecting against salmonid
CC rickettsial septicemia (SR) and other rickettsial diseases. The present
CC sequence represents an immunogenic epitope of the *P. salmonis* Ospa
XX protein. The peptide is used to raise rabbit anti-Ospa antibodies.
SQ Sequence 20 AA;

Query Match 13.4%; Score 112; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.9e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 110 PVRTRYRNKKORRQOYCRE 129
Db 1 PVRTRYRNKKORRQOYCRE 20

RESULT 6

ID AAB20105 standard; Protein; 224 AA.
XX AAB20105;

XX 23-APR-2001 (first entry)

DE Moraxella catarrhalis BASB113 protein.

KW BASB113: infection; otitis media; pneumonia; therapy; diagnosis;
XX antibacterial; antimicrobial.

OS Moraxella catarrhalis.

PN WO200100836-A1.

XX 04-JAN-2001.

PD 23-JUN-2000; 2000WO-EP05851.

XX 25-JUN-1999; 99GB-0015044.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Thonard J;

XX WPI: 2001-112458/12.

DR N-PSDB; AAF30043.

XX New BASB113 polypeptide isolated from Moraxella catarrhalis bacterium,
PT useful for diagnosing and producing vaccines against bacterial
PT infections such as otitis media and pneumonia -
PS Claim 1; Page 67; 86pp; English.

XX The present sequence is that of BASB113 protein from Moraxella
CC catarrhalis strain Mc931 (ATCC 43617), a causative agent of otitis
CC media in children and pneumonia in adults. The invention provides
CC BASB113 polypeptides, and polynucleotides encoding them, as well as
CC expression vectors, host cells and methods for producing them, as well as
CC polypeptides using recombinant methods. Also claimed is a vaccine
XX composition comprising a BASB113 polypeptide, an immunogenic

CC fragment of a BASB113 polypeptide, or a polypeptide having at least
CC 85% amino acid sequence identity to BASB113, or comprising a
CC polynucleotide encoding such a polypeptide. A claimed method of
CC diagnosing a Moraxella infection involves identifying a BASB113
CC polypeptide or antibody. A claimed therapeutic identifying a BASB113
CC in treating humans with M. catarrhalis infection comprises useful
CC 1 antibody directed against a BASB113 polypeptide. BASB113
CC polypeptides also have utility in raising specific antibodies.
XX and in screening for antibacterial drugs.

SQ Sequence 224 AA;

Query Match 12.9%; Score 108; DB 22; Length 224;
Best Local Similarity 32.4%; Pred. No. 0.0028;
Matches 36; Conservative 16; Mismatches 35; Indels 24; Gaps 5;

OY 8 GSSLITIVFLVGCANFRQEVGATG-----AVGCVAGQLFGKSGRVAMATG 59
Db 7 GVLLASSMALAGCANTGT--TGNGTFGCVANVKAIVGAVAGAL--GGRATSKATGG 60
OY 60 -----AVLGGIGSKTGQSDQDDKIKLNQSLKVKAGOVTRMRNPDGTN 104
Db 61 EKTGRDAILGAAGVAGAGAYMERQAK-----QIEQWQGTGVTHDIDTGN 107

RESULT 7

ID AAY34487 standard; Protein; 223 AA.
XX AAY34487;

XX 25-AUG-1999 (first entry)

DE Porphyromonas gingivalis protein PG3.

KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
XX vaccine; antigenic.

OS Porphyromonas gingivalis.

PN WO9929870-A1.

XX 17-JUN-1999.

PD 10-DEC-1998; 98WO-AU01023.

XX 04-AUG-1998; 98AU-0005028.

XX 10-DEC-1997; 97AU-0000839.

XX 31-DEC-1997; 97AU-0001182.

XX 30-JUN-1998; 98AU-0001546.

XX 10-MAR-1998; 98AU-0002264.

XX 09-APR-1998; 98AU-0002911.

XX 23-APR-1998; 98AU-0003128.

XX 05-MAY-1998; 98AU-0003338.

XX 22-MAY-1998; 98AU-0003654.

XX 29-JUL-1998; 98AU-0004917.

PA (CSL-) CSL LTD.

XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;

XX WPI: 1999-385613/32.

DR N-PSDB; AAX91705.

XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis

PS Claim 1; Page 469; 588pp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34593. AAX91802 to AAX91989 represent PCR primers used in the

Db

62 VCTLGGGALGIGASTIGGGRGQALAAVGAIGALGSKIEKMSQVNGAEI 114

RESULT 10

ABG15906

ID ABG15906 standard; Protein; 309 AA.

AC ABG15906;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #15897.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

XX N-PSDB; AAS80093.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT biodiversity for genetic disorders or other traits and to assess

PS Claim 20; SEQ ID No 46265; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences. ABG00010-ABG30377 represent novel human
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pot_sequences.

XX Sequence 309 AA;

Query Match

Best Local Similarity 11.7%; Score 98; DB 22; Length 309;

Matches 33; Conservative 11; Mismatches 39; Indels 14; Gaps 3;

QY 13 TTSVFLVGAQNSR-----QEVGATGAVVGVAGOLFEGK--SGVMAAIGAVLGL 65

pb 185 IISILPAKVAVNSQNRMAQAFGLIGAVGVIGHNVSGSNGTTRAGAVGAVGAA 244
J

QY 66 TSKI-----GQMDQODKIKLNLSLEKAKGQVT 95
Db 245 AGSMVNDKTLMEGVSLTYRKGTQVTSYOVGKEEFT 281

RESULT 11

AAB82611

ID AAB82611 standard; Protein; 528 AA.

AC AAB82611;

DT 02-OCT-2001 (first entry)

DE Spider recombinant silk protein PETNCDs.

KW Spider; orb-weaver; silk protein; PETNCDs; structural protein;

XX purification; fibre; spinning.

OS Nephila clavipes.

FH Key Location/Qualifiers

FT Misc-difference 417 /note= "encoded by TAT"

FT Misc-difference 427 /note= "encoded by CGA"

FT Misc-difference 522 /note= "encoded by GAG"

PN WO200153333-A1.

PD 26-JUL-2001.

PE 01-NOV-2000; 2000WO-US30086.

PR 20-JAN-2000; 2000US-0490291.

XX (MELL/) MELLO C M.

PA (ARCI/) ARCIDIAONO S.

PA (BUTL/) BUTLER M M.

PA (USAA) US SEC OF ARMY.

PI Mello CM, Arcidiacono S, Butler MM;

DR WPI; 2001-483136/52.

XX N-PSDB; AAB26304.

PT Recovering structural polypeptides in a biological sample, useful for
PT purifying and spinning spider silks and other structural proteins,
PT comprises treating the sample containing the polypeptides with an acid

PS Claim 2; Page 41-42; 49pp; English.

CC The present sequence is that of the orb-weaver spider (Nephila
CC clavipes) recombinant silk protein PETNCDs. The invention
CC provides methods for purifying and spinning spider silks and other
CC structural proteins. Organic acids are used to lyse recombinant
CC cells or other biological samples (such as non-recombinant
CC derived cells), and enrich the purity and yields of structural
CC proteins by hydrolysing many of the macromolecules while leaving
CC the structural proteins intact. In the case of silk proteins,
CC resulting lysate is further purified by ion-exchange or affinity
CC chromatography and processed into an aqueous-based mixture for
CC fibre spinning. In the present case, the PETNCDs gene was cloned
CC into vector pET24 for recombinant expression in Escherichia coli,
CC containing denatuant (3 M guanidine-HCl) and by affinity
CC chromatography on nickel-NTA agarose. Products obtained using the
CC methods of the invention can be used in the construction of many
CC ballistics including films, fibres, woven articles, sutures,
CC method has the following advantages over prior art: it involves
CC fewer steps, requires less time and smaller volumes of reagents,

QY 126 YCREFOQKAMIAQKQ 141
 Db 582 QQQQQQQQQHQHQQQQQQ 597

RESULT 14

AAV40097
 ID AAV40097 standard; protein: 651 AA.

AC AAV40097;

DT 19-NOV-1999 (first entry)

DE Spider silk protein spidroine major 1.

KW Spider silk protein; spidroine major 1; cosmetic; make-up;

KW dermatological compositions; hair care; skin care; sunscreen;

OS Nephila clavipes.

PN FR2774588-A1.

PD 13-AUG-1999.

PF 11-FEB-1998; 98FR-0001614.

PR 11-FEB-1998; 98FR-0001614.

PA (OREA) L/OREAL SA.

PI Philippe M, Garson JC, Arraudeau JP;

DR WPI: 1999-510729/43.

PT Cosmetic or dermatological composition containing spider silk protein,
 for hair or skin care, in make-up or sunscreens

PS Claim 3; Fig 1: 32pp; French.

CC The present sequence represents the natural spider silk protein
 CC spidroine major 1. The protein improves the moisturizing/softening
 CC action of the compositions. The protein, and its fragments are used
 CC in cosmetic or dermatological compositions. These compositions have
 CC use as hair or skin care products; and make-up or sunscreens.
 CC As the protein is a good, persistent film-formers on the skin
 CC of low surface density, it can be used for delivery of active
 CC agents that are generally difficult to administer, e.g., vitamins,
 CC hormones, moisturizers or agents for treating disorders of the
 CC skin and hair.

SQ Sequence 651 AA;

Query Match 9.7%; Score 81.5; DB 20; Length 651;
 Best Local Similarity 36.8%; Pred. No. 6.7;
 Matches 25; Conservative 3; Mismatches 17; Indels 23; Gaps 2;

QY 28 QEVGAATGAVVG---GVAGQ-----LFGKSGRVAMAGAVLGG 64
 Db 491 QGAGAAAAAAVAGAGQEBIRGQAGAGGYGGLGSGGRGGLGGGAGAAAAAGAGAGGG 550
 QY 65 LIGSKTGO 72
 Db 551 LGGGAGAGQ 558

RESULT 15

ID AARI4308
 XX AARI4308 standard; Protein: 718 AA.

AC AARI4308;

DT 15-JAN-1992 (first entry)

XX N.clavipes dragline silk protein-1.
 DE protein superfibre; major ampullate silk; orb web spider.

KW Nephila clavipes.

PN EP452925-A.

PD 23-OCT-1991.

PF 18-APR-1991; 91EP-0106217.

PR 20-APR-1990; 90US-0511792.

PA (UYWY-) UNIV OF WYOMING.

PI Lewis RV, Xu M, Hlman M;

DR WPI: 1991-312199/43.

PT DNA encoding spider silk protein-1 and 2 and variants - isolated
 from Nephila clavipes, for prodn. of spider silk protein and
 fibres having desired characteristics

PS Claim 15; Page 23; 48pp; English.

CC The spider silk protein contains a basic 34 amino acid repeat. The
 CC repeat itself contains 3 regions. The first comprises 0-9 amino
 CC acids with a sequence AGR(GGX)2. This region is not highly
 CC conserved. The second region has a sequence GAG(A)x which is highly
 CC and is 15 amino acids long. The third segment is (GGX)5
 CC cases X is A/Q, Y or L. Removal of the poly-(Ala) segments results
 CC in a silk having lower elasticity.

SQ Sequence 718 AA;

Query Match 9.7%; Score 81.5; DB 12; Length 718;
 Best Local Similarity 36.8%; Pred. No. 7.6;
 Matches 25; Conservative 3; Mismatches 17; Indels 23; Gaps 2;

QY 28 QEVGAATGAVVG---GVAGQ-----LFGKSGRVAMAGAVLGG 64
 Db 491 QGAGAAAAAAVAGAGQEBIRGQAGAGGYGGLGSGGRGGLGGGAGAAAAAGAGAGGG 550
 QY 65 LIGSKTGO 72
 Db 551 LGGGAGAGQ 558

Search completed: October 27, 2002, 10:57:54
 Job time : 23.8238 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 11:00:01 ; Search time 18.1865 Seconds

(without alignments)
1340.986 Million cell updates/sec

Title: US-09-677-374-2

Sequence: 1 MNRGCLQGSSLLIISVFLVG.....IYGTACRQPDGRMOWISTEK 162

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_ricent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvrius:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	836	100.0	162	2	09F9K8
2	313.5	37.5	148	2	054381
3	291	34.8	159	2	09F9F2
4	266.5	31.9	137	2	052252
5	260.5	31.2	144	2	09K2N6
6	259.5	31.0	137	2	031065
7	259.5	31.0	144	2	09K4W8
8	252	30.1	154	2	053154
9	247	29.5	151	2	09F909
10	237.5	28.4	131	2	09F001
11	237.5	28.4	131	2	052637
12	236.5	28.3	131	2	09L522
13	223.5	26.7	105	2	031208
14	143	17.1	199	16	0985G4
15	140	16.7	77	2	09AGC7
16	130.5	15.6	136	16	092R89

17	128.5	15.4	182	16	09HXT3	09hxt3 pseudomonas
18	116	13.9	154	16	09I4S1	09i4s1 pseudomonas
19	112	13.4	257	16	09PGX0	09pgx0 xylella fas
20	111	13.3	139	2	052854	052854 rhizobium 1
21	110.5	13.2	232	16	09A3X8	09a3x8 caulobacter
22	108.5	13.0	155	2	09F6B1	09f6b1 edwardsiell
23	107.5	12.9	153	2	069776	069776 rhizobium e
24	107.5	12.9	155	2	09RA95	09ra95 serratia sp
25	106.5	12.7	221	16	092ST9	092st9 rhizobium m
26	106	12.7	155	2	09RB08	09rb08 pectobacter
27	105.5	12.6	223	2	09XC44	09xca4 porphyromon
28	101.5	12.1	304	16	09I762	09i762 pseudomonas
29	99.5	11.9	83	16	092LP2	092lp2 rhizobium m
30	99	11.8	154	16	09CNR3	09cnc3 pasteurilla
31	98	11.7	172	16	P76572	P76572 escherichia
32	98	11.7	220	2	09YAX8	09yax8 pseudomonas
33	96	11.5	223	16	09KSR1	09ksr1 vibrio chol
34	95.5	11.4	79	16	092NT4	092nt4 rhizobium m
35	94.5	11.3	838	2	09AL49	09al49 shigella fl
36	94	11.2	608	10	09SUX1	09sux1 arabidopsis
37	93.5	11.2	257	16	09A8M8	09a8m8 caulobacter
38	92	11.0	105	16	098P93	098p93 rhizobium 1
39	91	10.9	544	5	046171	046171 nephila cla
40	90	10.8	105	16	0983Y0	0983y0 rhizobium 1
41	89	10.6	137	16	09HU07	09hu07 pseudomonas
42	88.5	10.6	161	2	068681	068681 bacillus me
43	88.5	10.6	216	16	098NA9	098na9 rhizobium 1
44	87.5	10.5	407	16	031557	031557 bacillus su
45	87	10.4	201	16	0930X4	0930x4 rhizobium m

ALIGNMENTS

RESULT 1
ID 09F9K8 PRELIMINARY: PRT: 162 AA.

AC 09F9K8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE 17 KDA ANTIGEN.
GN OSPA.
OS Piscirickettsia salmonis.
OC Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;
OC Piscirickettsia.
OX NCBI_TaxID=1238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LF-89;
RA Kuzyk M.A., Thornton J.C., Kay W.W.;
RT Identification of a genus-common Rickettsial surface antigen in the
RT salmonid pathogen Piscirickettsia salmonis.*;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF184152; ANG17000.1; -
SQ SEQUENCE 162 AA; 17661 MW; DDE39BEFD94A527E CRC64;

Query Match 100.0%; Score 836; DB 2; Length 162;

Best local similarity 100.0%; Pred. No. 9.6e-68;

Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MNRGCLQGSSLLIISVFLVGCACNPFSSROEYGAATGAVGCVAGQLFGKSGRYAMAIGCA	60
DB	1	MNRGCLQGSSLLIISVFLVGCACNPFSSROEYGAATGAVGCVAGQLFGKSGRYAMAIGCA	60
QY	61	VLGGLGSKTGQSMDOODKTKLNQSLSEKYKAGGVTFMRNPDGTSNSVSEPVRTYQYRNO	120
DB	61	VLGGLGSKTGQSMDOODKTKLNQSLSEKYKAGGVTFMRNPDGTSNSVSEPVRTYQYRNO	120
QY	121	ERRQOYCRFEFOKRAMIAGOKOEIYGTACRQPDGRMOWISTEK	162
DB	121	ERRQOYCRFEFOKRAMIAGOKOEIYGTACRQPDGRMOWISTEK	162

QY 64 GLIGSKTIGQSDMDQDK---IKTNSLEKVGKAGVTRMRNPDGNSYSVEPVRTYGRYNK 119
 Db 65 AVLGSGIGAGNDEEDQRRLAELTSGRALERTPGSTGVEMWRNPDNCHGVTPNKTY----- 119
 QY 120 QERRQOYCRFEQKAMIAAGKKEITGDRCPDRGMYIS 159
 Db 120 RNSTGYCREYTOTVTYIGGKQKAYGNACROPDGIWQVNV 159

 RESULT 4
 052252
 ID 052252 PRELIMINARY: PRT: 137 AA.
 AC 052252:
 DT 01-JUN-1998 (TREMBLrel, 06, Created)
 DT 01-JUN-1998 (TREMBLrel, 06, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel, 19, Last annotation update)
 OS 17 KDA ANTIGEN (FRAGMENT).
 OS Rickettsia cooley.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxId=69410;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Billings A.N., Teltow G.J., Walker D.H.;
 RT "Molecular characterization of a novel spotted fever group rickettsial
 RT species from Ixodes scapularis in Texas."
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF031534; AAB95267.1; --
 FT NON_TER 1
 SQ SEQUENCE 137 AA; 14215 MW; A27597A9AFD85FC3 CMC64;
 Query Match 31.9%; Score 266.5; DB 2: Length 137;
 Best Local Similarity 39.6%; Pred. No. 14e-16;
 Matches 53; Conservative 23; Mismatches 49; Indels 9; Gaps

 QY 23 QNFSRQEGVATAVVGVAGVOLFEGKSSRVMAIGAVLGLGSKTIGSDMDQDK--- 79
 Db 7 RGNMKQGTGTLTGAGGALLGSGFGRGGLVGVGVGLGAVLGIGAGMDEQDRRLA 66
 QY 80 -IKTNSLEKVGKAGVTRMRNPDGNSYSVEPVRTYGRYNKQERRQOYCRFEQKAMIA 138
 Db 67 ELMSGVALMAASGSGSTWRNPDNMGVTPNKTY-----RNSTGYCREYTOTVTYIG 121
 QY 139 QKEITGACROPD 152
 Db 122 KQKAYGNACROPD 135

 RESULT 5
 09K2N6
 ID 09K2N6 PRELIMINARY: PRT: 144 AA.
 AC 09K2N6:
 DT 01-OCT-2000 (TREMBLrel, 15, Created)
 DT 01-OCT-2000 (TREMBLrel, 15, Last sequence update)
 DE 17KDA ANTIGEN (17 KDA ANTIGEN) (FRAGMENT).
 OS male-killing Rickettsia from Adalia bipunctata.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxId=38028;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Schlenker H.J.G.V.D., Habig M., Sloggett J.J., Webberley M.K.,
 RA Bertrand D., Hurst G.D.D., Majerus M.E.N.;
 RT "On the evolution of male-killing Monophyletic origin and horizontal
 RT transfer of male-killing Rickettsia (a-Proteobacteria) from two con-
 RL (Coleoptera: Coccinellidae)."
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ269518; CAB96383.1; --
 FT EMBL; AJ269517; CAB96382.1; --
 FT NON_TER 1
 SQ SEQUENCE 144 AA; 14415 MW; A27597A9AFD85FC3 CMC64;
 Query Match 31.9%; Score 266.5; DB 2: Length 137;
 Best Local Similarity 39.6%; Pred. No. 14e-16;
 Matches 53; Conservative 23; Mismatches 49; Indels 9; Gaps

SO	SEQUENCE	144 AA;	14785 MW;	C6254739CCAS6AE7	CRC64;
QY	Query Match	31.2%;	Score 260.5;	DB 2;	Length 144;
Db	Best Local Similarity	39.7%;	Pred. No. 5,1e-16;		
Matches	52;	Conservative	22;	Mismatches	48; Indels 9; Gaps 2;
QY	26 SROEFGAATGAVGVAGAGLFEKSGGRAAMAIGAVLGGLTSKIGOSDODDK----	IK 81			
Db	17 NKOGTGLLGGAGGALLTSQFEGKGGOLVGVGVALLGAVLGQIGAGDEDDRRLAE	LT 76			
QY	82 LNSGTEKXKAGOVTFWRNPDDTGNSTSVPEPRTYGRYKNGEERQOYCREFOQKAM	IAGOKO 141			
Db	77 SORALEAAPSGSNVEMRNPDMNGNHYVTIPNKTY-----	RNSTGQYCREYTORVIVIGKQO 131			
QY	142 EIVGTACRQPD 152				
Db	132 KAYGNACRQPD 142				
RESULT 6					
ID	O31065	PRELIMINARY;	PRT;	137 AA.	
AC	O31065; O9M02;				
DT	01-JAN-1998 (TREMBLrel. 05, Created)				
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)				
DE	01-OCT-2001 (TREMBLrel. 18, Last annotation update)				
DE	17 KDA ANTIGEN (17 KDA PROTEIN) (FRAGMENT).				
OS	Rickettsia honei.				
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;				
OC	Rickettsiaceae; Rickettsiaseae; Rickettsia.				
OX	NCBI_TaxID=37816;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RP	STRAIN-TT-118;				
RA	Billings A.N., Yu X.-J., Teel P.D., Walker D.H.;				
RL	Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=99045882; PubMed=9828442;				
RA	Stenos J., Roux V., Walker D., Raoul D.;				
RT	"Rickettsia honei sp. nov., the aetiological agent of Flinders Island				
RT	spotted fever in Australia."				
DR	Int. J. Syst. Bacteriol. 46:1399-1404(1998).				
DR	EMBL; AF027124; AAB81846.1; -				
DR	EMBL; AF060706; AAD20231.1; -				
DR	EMBL; AF060704; AAD20230.1; -				
FT	NON_TER	1			
SO	SEQUENCE	137 AA;	14167 MW;	75ECID0D745B428C	CRC64;
QY	Query Match	31.0%;	Score 259.5;	DB 2;	Length 137;
Db	Best Local Similarity	39.7%;	Pred. No. 5,9e-16;		
Matches	52;	Conservative	22;	Mismatches	48; Indels 9; Gaps 2;
QY	26 SROEFGAATGAVGVAGAGLFEKSGGRAAMAIGAVLGGLTSKIGOSDODDK----	IK 81			
Db	10 NKOGTGLLGGAGGALLTSQFEGKGGOLVGVGVALLGAVLGQIGAGDEDDRRLAE	LT 69			
QY	82 LNSGTEKXKAGOVTFWRNPDDTGNSTSVPEPRTYGRYKNGEERQOYCREFOQKAM	IAGOKO 141			
Db	70 SORALEETAPSGSNVEMRNPDMNGNHYVTIPNKTY-----	RNSTGQYCREYTORVIVIGKQO 124			
QY	142 EIVGTACRQPD 152				
Db	125 KAYGNACRQPD 135				
RESULT 7					
ID	O9K4W8	PRELIMINARY;	PRT;	144 AA.	
AC	O9K4W8;				
DT	01-OCT-2000 (TREMBLrel. 15, Created)				
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				

Query Match	Best Local Similarity	Score	DB 2	Length	Indels	Gaps
Matches 52; Conservative 22; Mismatches 48; Indels 9; Gaps 2;	31.0%;	259.5;	DB 2;	144;		
Best Local Similarity 39.7%; Pred. No. 6.3e-16;						
Matches 52; Conservative 22; Mismatches 48; Indels 9; Gaps 2;						
17 KDA ANTIGEN (FRAGMENT).						
male-killing Rickettsia from Adalia decempunctata.						
Bacteria: Proteobacteria; alpha subdivision; Rickettsiales;						
Rickettsiaceae; Rickettsiaseae; Rickettsia.						
NCBI_TaxID=120393;						
SEQUENCE FROM N.A.						
MEDLINE=20575219; PubMed=11133455;						
Schulenburg H.J.G.V.D., Habig M., Sloggett J.J., Webberley M.K.,						
Bertrand D., Hurst G.D., Majerus M.E.N.;						
"Incidence of male-killing Rickettsia spp. (alpha-Proteobacteria) in						
the Ten-Spot Ladybird Beetle Adalia decempunctata L. (Coleoptera:						
Coccinellidae)."?						
Appl. Environ. Microbiol. 67:270-277(2001).						
EMBL; AJ269516; CAB96381.1; -.						
NON TER 1 144						
NON TER 1 144						
SEQUENCE 144 AA; 14801 MW; C025472F16A56AE7 CRC64;						
Query Match	31.0%;	Score 259.5;	DB 2;	Length 144;		
Best Local Similarity 39.7%; Pred. No. 6.3e-16;						
Matches 52; Conservative 22; Mismatches 48; Indels 9; Gaps 2;						
26 SROEAGATGAVGVGAQLPEKSGSRVAMALGAVLGLGISKTIQSDODK---IK 81						
17 NKQGTGTLGGAGGALLSQPEKGKGLVGVGALLGAVLGGQIAGDEDDRLAELT 76						
82 LNQSLEKVGAVTWRMPNDPTGNSYSEVPRTYRNKQERQOQCREFOKAMTAGOKO 141						
77 SORALEAPSGSNVWRPNPDNGNHGVTPNKTYY-----RNSTGYCREYTOYVIGKQ 131						
142 ELYGTACROPD 152						
132 KSYGNACROPD 142						
RESULT 8						
053154 PRELIMINARY; PRT; 154 AA.						
053154 AC						
01-NOV-1996 (TREMBLrel. 01, Created)						
01-NOV-1996 (TREMBLrel. 01, Last sequence update)						
01-NOV-1998 (TREMBLrel. 08, Last annotation update)						
(CLONE PRB F15F 1), 5' END CDS (FRAGMENT).						
Rickettsia sp.						
Bacteria: Proteobacteria; alpha subdivision; Rickettsiales;						
Rickettsiaceae; Rickettsiaseae; Rickettsia.						
NCBI_TaxID=789;						
SEQUENCE FROM N.A.						
MEDLINE=93084757; PubMed=1452660;						
Baird R.W., Lloyd M., Stenos J., Ross B.C., Stewart R.S., Dwyer B.;						
"Characterization and comparison of Australian human spotted fever						
group Rickettsiae."						
J. Clin. Microbiol. 30:2896-2902(1992).						
EMBL; M99391; AAA73386.1; -.						
NON TER 154						
NON TER 154						
SEQUENCE 154 AA; 15849 MW; F5C35855EDBA439D2 CRC64;						
Query Match	30.1%;	Score 252;	DB 2;	Length 154;		
Best Local Similarity 35.5%; Pred. No. 3.3e-15;						
Matches 55; Conservative 29; Mismatches 57; Indels 14; Gaps 4;						
9 SLLIIISV---FLVGC--AQNFSROEVGAATGAVGVGAQLPEKSGSRVAMALGAVLG 63						
5 SKIMIIALTSMQACNCPGCKNNKQGTGTLGAGAGALLGSGFGKKGGLVGVGALLG 64						
64 GLIGSGISGSDODK---IKLNSLEKVGAVTWRMPNDPTGNSYSEVPRTYORYNK 119						
65 AVLGGQIAGMDEQDRLAELTSGRALETAPSGSNVWRPNPDNGSYGVTPNKTYYRNSG 124						
120 QERROOYCREFOOKAMIAQOKEIYGTACROPDGR 154						

Db 125 QD-----CRVYTOTVIVIGKQKAKYGNACRQPDGQ 154

RESULT 9

Q9F909

AC Q9F909 PRELIMINARY; PRT; 151 AA.

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DE OUTER MEMBRANE PROTEIN (FRAGMENT)

OS Rickettsia helvetica

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OX Rickettsiaceae; Rickettsiidae; Rickettsia

NCBI_TaxID=35789;

RP SEQUENCE FROM N.A.

RA Nilsson K., Paulson C.;

RT "Novel peptide diagnostic reagent and kit for detection of Rickettsiosis."

DR Submitted (AUG-1999) to the EMBL/Genbank/DBJ databases.

FT EMBL; AF181036; AAC09427.1; -

NON_TER 151

SEQUENCE 151 AA; 15621 MW; B77407B9C71E4B39 CRC64;

Query Match

Best Local Similarity 29.5%; Score 247; DB 2; Length 151;

Matches 53; Conservative 26; Mismatches 57; Indels 14; Gaps 4;

QY 9 SSLIIVSV---FLVGC--AQNRSREVGATGAVGVAGQLFGKSGRVAMATGAVIG 63

Db 5 SKIMITALAASMLQACNGPGKMGKGTGTLGGAGALLSGQKGGQGLVGVGALLG 64

QY 64 GLIGSKIGQSDQDK---IKLNSLEKVKAGQVTRWRNPDTGNSYSVEPVRTYQRYNK 119

Db 65 AVLGQIYAGDEDDRLAELTSGRALEAPSGSNVEMRPDNGYVTPNKTY----- 119

QY 120 QERRQCYCREFQOKAMTAGOKOELTYGACROP 151

Db 120 RNSTGYCYCREYTOTVIVIGKQKAKYGNACRQPDGQ 151

RESULT 10

Q9F001

Q9F001

AC Q9F001 PRELIMINARY; PRT; 131 AA.

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DE 17 KDA PROTEIN (FRAGMENT)

OS Rickettsia sp. California 2.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OX Rickettsiaceae; Rickettsiidae; Rickettsia

NCBI_TaxID=147259;

RP SEQUENCE FROM N.A.

RA Raoult D.;

RT "A new SFG rickettsia isolated from fleas."

DR Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.

FT EMBL; AF210693; AAG48554.1; -

NON_TER 131

SEQUENCE 131 AA; 13374 MW; 23C8819B29FF660 CRC64;

Query Match

Best Local Similarity 28.4%; Score 237.5; DB 2; Length 131;

Matches 48; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

QY 26 SROEVGATGAVGVAGQLFGKSGRVAMATGAVLGLISKIGQSDQDK-----IK 81

Db 10 NKQGTGTLGGAGALLGSGQFGKGGQGLVGVGALLAVLGGQIGAGMDQDRRLAELT 69

QY 82 LNSLEKVKAGQVTRWRNPDTGNSYSVEPVRTYQRYNKQERRQCYCREFQOKAMTAGOKO 141

Db 70 SORALEATPSGTSVEMRPDNGHGYVTPNKTY-----RNSTGYCYCREYTOTVIVIGKQO 124

QY 142 ELYGTAC 148

Db 125 KAYGNAC 131

RESULT 11

Q52637

Q52637

AC Q52637 PRELIMINARY; PRT; 131 AA.

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DE 17 KDA ANTIGEN (FRAGMENT)

OS Rickettsia sp.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OX Rickettsiaceae; Rickettsiidae; Rickettsia

NCBI_TaxID=789;

RP SEQUENCE FROM N.A.

RA Warren J.H., Hurst G.D., Zhang W., Breuer J.A., Stouthamer R.,

RT "Rickettsial relative associated with male killing in the ladybird

beetle (Adalia bipunctata)."

DR EMBL; U04162; AAA19235.1; -

FT NON_TER 131

SEQUENCE 131 AA; 13344 MW; A1DCF71050DF52DF CRC64;

Query Match

Best Local Similarity 28.4%; Score 237.5; DB 2; Length 131;

Matches 48; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

QY 26 SROEVGATGAVGVAGQLFGKSGRVAMATGAVLGLISKIGQSDQDK-----IK 81

Db 10 NKQGTGTLGGAGALLGSGQFGKGGQGLVGVGALLAVLGGQIGAGMDQDRRLAELT 69

QY 82 LNSLEKVKAGQVTRWRNPDTGNSYSVEPVRTYQRYNKQERRQCYCREFQOKAMTAGOKO 141

Db 70 SORALEATPSGTSVEMRPDNGHGYVTPNKTY-----RNSTGYCYCREYTOTVIVIGKQO 124

QY 142 ELYGTAC 148

Db 125 KAYGNAC 131

RESULT 12

Q9L522

Q9L522

AC Q9L522 PRELIMINARY; PRT; 131 AA.

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DE 17 KDA SURFACE ANTIGEN (FRAGMENT)

OS Rickettsia peacockii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OX Rickettsiaceae; Rickettsiidae; Rickettsia

NCBI_TaxID=47589;

RP SEQUENCE FROM N.A.

RA Simer J.A., Palmer A.T., Munderloh U.G., Kurtli T.J.;

RT "Isolation of a spotted fever group rickettsia, Rickettsia peacockii,
in a Rocky Mountain wood tick, Dermacentor andersoni, cell line."
RT Appl. Environ. Microbiol. 67:546-552(2001).
DR EMBL: AF260571; AAF69012.1; -.

FT NON_TER 131 131
SQ SEQUENCE 131 AA; 13413 MW; 228C020550CA9D0 CRC64;

Query Match 28.3%; Score 236.5; DB 2; Length 131;
Best Local Similarity 37.8%; Pred. No. 6.6e-14;
Matches 48; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

QY 26 SROEVGAATGAVGVAGVAGLFGKSGRVAMAIGGAVGLGSKIGSQMDQDK----IK 81
DB 10 NKGSTGLGAGAGALLGSGFGKSGKGLGVGVGALLGAVGIGGIGAGMDQDRLELT 69
QY 82 LNSLEKVKAGQVTRMRNPDTGNSYSVEPVTRQYRKQERRROOYCFEFOOKAMIAOKO 141
DB 70 SORALETPAGSNNVEMPNNGVYTPNKTY-----RNSTGQYCREYTQTVVIGGKQO 124
QY 142 EITGTAC 148
DB 125 KAIGNAC 131

RESULT 13
031208 PRELIMINARY; PRT; 105 AA.

AC 031208;
DT 01-JUN-1998 (TREMblrel. 05, Created)
DT 01-JUN-1998 (TREMblrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE 17 KDA ANTIGEN (FRAGMENT).
OS Rickettsia sp. "La Copita".
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=69475;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LA COPITA;
RA MEDLINE=98367252; PubMed=9701930;
RA Ballings A.N., Yu X.J., Teel P.D., Walker D.H.;
RT "Detection of a spotted fever group rickettsia in Amblyomma cajennense
[Acari: Ixodidae] in south Texas".
RL J. Med. Entomol. 35:474-478(1998).
DR EMBL: AF03499; AAB86943.1; -.
FT NON_TER 105 105
SQ SEQUENCE 105 AA; 11236 MW; 7BDD35104701D4FB CRC64;

Query Match 26.7%; Score 223.5; DB 2; Length 105;
Best Local Similarity 41.2%; Pred. No. 7.4e-13;
Matches 47; Conservative 17; Mismatches 33; Indels 17; Gaps 3;

QY 43 GOLFSGSGRVAMAIGGAVGLGSKIGSQMDQDK----IKLNSLEKVKAGQVTRMR 98
DB 3 GOLVGAV-----GALLGAVIGGIGAGMDQDRLELTQSRLETPAGSSNNVEMR 54
QY 99 NPDTGNSYSVEPVTRQYRKQERRROOYCFEFOOKAMIAOKOIEYGTACRQPD 152
DB 55 NPDGNGVYTPNKTY-----RNSTGQYCREYTQTVVIGGKQAKAYGNACRQPD 103

RESULT 14
0985G4 PRELIMINARY; PRT; 199 AA.
AC 0985G4;
DT 01-OCT-2001 (TREMblrel. 18, Created)
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)
DE MUR7687 PROTEIN.
GN MUR7687.
OS Rhizobium loti (Mesorhizobium loti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;

RP SEQUENCE FROM N.A.
RC STRAIN=MAFE303099;
RA MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Katsunabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsuno M., Matsuno A.,
RA Hoshida Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL: AF03012; BAB54098.1; -.
KW Complete proteome.
SQ SEQUENCE 199 AA; 20517 MW; 98E6F20A734637AA CRC64;

Query Match 17.1%; Score 143; DB 16; Length 199;
Best Local Similarity 31.8%; Pred. No. 2.9e-05;
Matches 35; Conservative 15; Mismatches 50; Indels 10; Gaps 3;

QY 50 SGRVAMAIGGAVGLGSKIGSQMDQDKIKL----NLSLEKVKAGQVTRMRNPDTGNS 105
DB 92 SGKVTKSLISAMDGGLGSGISGNLSDEKRSALAEYKALEYTTSGOKVAMKGDQASHY 151
QY 106 YSVEPVTRQYRKQERRROOYCFEFOOKAMIAOKOIEYGTACRQPGRW 155
DB 152 GEVVPAPY-RVGSQD-----CROYTQTVFTGAGVATARGACRMDGWSW 195

RESULT 15
09AGC7 PRELIMINARY; PRT; 77 AA.

AC 09AGC7;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE 17 KDA ANTIGEN (FRAGMENT).
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=785;

RN [1]
RP SEQUENCE FROM N.A.
RA Paspaliaris V., Liedtke B., Viletta L., Whiting J.L.;
RT "Rickettsia typhus and Rickettsia felis rickettsiases in Australia."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF336794; AAK21272.1; -.

FT NON_TER 77 77
SQ SEQUENCE 77 AA; 7955 MW; BIE447C037263918 CRC64;

Query Match 16.7%; Score 140; DB 2; Length 77;
Best Local Similarity 38.7%; Pred. No. 1.6e-05;
Matches 29; Conservative 13; Mismatches 29; Indels 4; Gaps 1;

QY 43 GOLFSGSGRVAMAIGGAVGLGSKIGSQMDQDK----IKLNSLEKVKAGQVTRMR 98
DB 3 GSGFGHCKGGLGVGVGALLGAVIGGIGAGMDQDRLELTQSRALLESAPSSNNVEMR 62
QY 99 NPDTGNSYSVEPVTR 113
DB 63 NPDGNGVYTPNKTY 77

Search completed: October 27, 2002, 11:10:22
Job time : 19.1865 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 11:00:36 : Search time 6.15544 Seconds

(without alignments)
1019.028 Million cell updates/sec

Title: US-09-677-374-2

Sequence: 1 MNRGCLQSSLIITISVFLVG.....IYGTACRQDPGRMQVISTEK 162

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	236	35.4	159	17KD_RICPR	P16624 rickettsia
2	234	35.2	159	17KD_RICJA	O52764 rickettsia
3	293	35.0	159	17KD_RICCN	P05372 rickettsia
4	284.5	34.0	159	17KD_RICTY	P22882 rickettsia
5	276	33.0	154	17KD_RICAU	P50928 rickettsia
6	263	31.5	154	17KD_RICAM	P50927 rickettsia
7	261	31.2	154	17KD_RICPA	P50931 rickettsia
8	261	31.2	154	17KD_RICRH	P50929 rickettsia
9	258	30.9	154	17KD_RICMO	P29687 rickettsia
10	157	18.8	80	17KD_RICCA	P31484 yersinia en
11	113.5	13.6	155	PCP_YEREN	P55741 yersinia en
12	105.5	12.6	155	SLVB_ECOLI	O53549 salmonella
13	105.5	12.6	155	SLVB_SALT	P10335 haemophilus
14	102	12.2	155	PCP_HAEIN	P37796 escherichia
15	99	11.8	179	YCFJ_ECOLI	P06384 bos taurus
16	90	10.8	526	KICJ_BOVIN	P54334 bacillus su
17	90	10.8	1332	XKDO_BACSU	P45931 bacillus su
18	89	10.6	1585	YOBO_BACSU	P37723 salmonella
19	87.5	10.5	72	OSMB_SALT	P17853 salmonella
20	86.5	10.3	72	OSMB_ECOLI	P17853 salmonella
21	84.5	10.1	243	CYSH_SALT	O92188 rattus norv
22	84.5	10.1	301	STXG_RAT	O00310 candida alb
23	84.5	10.1	431	KRE2_CANAL	P33476 bluetongue
24	84.5	10.1	526	VP5_BTVAL	O66867 aquilex aeo
25	82	9.8	132	V615_AOUAE	P13135 bos taurus
26	82	9.8	263	CANS_BOVIN	P48837 saccharomyc
27	81.5	9.7	541	NU57_YEAST	P19837 nephila cla
28	81.5	9.7	747	SPD1_NEPCL	P46724 mycobacteri
29	80.5	9.6	467	HEM1_MYCLE	P34309 caenorhabdi
30	79.5	9.5	113	YKR3_CABEL	P06813 oryctolagus
31	79.5	9.5	266	CANS_RABIT	P19482 bos taurus
32	79.5	9.5	359	ATPA_BOVIN	P15999 rattus norv
33	79.5	9.5	543	ATPA_RAT	

34	79.5	9.5	553	1	ATPA_HUMAN	P25705 homo sapien
35	79	9.4	593	1	KICJ_HUMAN	P13645 homo sapien
36	78.5	9.4	553	1	ATPD_BOVIN	P19483 bos taurus
37	78.5	9.4	553	1	ATPA_MOUSE	O03285 mus musculus
38	78	9.3	865	1	VGLB_HSYMD	P18538 marek's dis
39	77.5	9.3	219	1	YIAD_ECOLI	P37665 escherichia
40	77	9.2	269	1	CANS_MOUSE	O88456 mus musculus
41	77	9.2	393	1	CSP_PLABR	P14593 plasmodium
42	77	9.2	429	1	CSP_PLAMA	P13815 plasmodium
43	77	9.2	506	1	ATPA_ANASP	P12405 anaobena sp
44	77	9.2	514	1	ATPA_THIFE	P41167 chloobacillu
45	76.5	9.2	243	1	CYSH_ECOLI	P17854 escherichia

ALIGNMENTS

RESULT 1
ID 17KD_RICPR STANDARD; PRT; 159 AA.

AC P16624;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 17 kDa surface antigen precursor.

GN OMP OR RP833.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX NCBI_Taxid=782;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E.
RA MEDLINE=69359171; Pubmed=2768201;
RX Anderson B.E., Tzianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen
RT gene.";
RT J. Bacteriol. 171:5199-5201(1989).

RL [2]
RN SEQUENCE FROM N.A.
RC STRAIN=MADRID E.
RA MEDLINE=99039499; Pubmed=9823893;
RX Anderson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RT Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; M28482; AAA26378.1; ALT_SEQ.
DR EMBL; AJ235273; CAA15258.1; -
DR PIR; D33971; D33971.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
SQ SEQUENCE 159 AA; 16672 MW; A33D404B55EBB071 CRC64;

Query Match 35.4%; Score 296; DB 1; Length 159;
Best Local Similarity 38.1%; Pred. No. 9.9e-19;
Matches 61; Conservative 30; Mismatches 55; Indels 14; Gaps 4;

```

OY 9 SLLIISV---FLVGC--AQNFSROEVGATGAVGVAGGOLFGRSGRYVMAAIGAVLG 63
DB 5 SKIMIIALASMLACNCGSGMANKOGTGLLGAGALLGSGFQGGGLVGVGALLG 64
OY 64 GLISGRKIGSGMDQDK---IKLNQSLKRVKAGVTRRPNDTGNSSYSEVPRTYQRYNK 119
DB 65 AVLGQIGASMDQDRLELTSQRALESASSNINRPNPDNGNYTPNKTY----- 119
OY 120 QERROOYCRFEQOKAMIAQKOEIYGNACROPDGRMOWIS 159
DB 120 RNSAGOCREYTYQVYIIGKQOKRYGNACROPDGRMOWYVN 159

RESULT 2
17KD_RICJA
ID 17KD_RICJA STANDARD; PRT: 159 AA.
AC 052764;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 17 kDa surface antigen precursor.
GN OMP.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
RX NCBI_TaxID=35790;
RP SEQUENCE FROM N.A.
RC STRAIN-YH.
RX MEDLINE=95229950; PubMed=7714214;
RA Furuya Y., Katayama T., Yoshida Y., Kaiho I.;
RT "Specific amplification of Rickettsia japonica DNA from clinical
RL J. Clin. Microbiol. 33:487-489(1995).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, D16515; BAA03965.1;
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KM Outer membrane: Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 159
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
SQ SEQUENCE 159 AA; 16554 MW; CDDCE7CEBDD641 CRC64;

Query Match 35.2%; Score 294; DB 1; Length 159;
Best Local Similarity 38.1%; Pred. No. 1.5e-18;
Matches 61; Conservative 30; Mismatches 55; Indels 14; Gaps 4;
OY 9 SLLIISV---FLVGC--AQNFSROEVGATGAVGVAGGOLFGRSGRYVMAAIGAVLG 63
DB 5 SKIMIIALASMLACNCGSGMANKOGTGLLGAGALLGSGFQGGGLVGVGALLG 64
OY 64 GLISGRKIGSGMDQDK---IKLNQSLKRVKAGVTRRPNDTGNSSYSEVPRTYQRYNK 119
DB 65 AVLGQIGASMDQDRLELTSQRALESASSNINRPNPDNGNYTPNKTY----- 119
OY 120 QERROOYCRFEQOKAMIAQKOEIYGNACROPDGRMOWIS 159
DB 120 RNSAGOCREYTYQVYIIGKQOKRYGNACROPDGRMOWYVN 159

RESULT 3
17KD_RICCN
ID 17KD_RICCN STANDARD; PRT: 159 AA.

```

```

AC P05372;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 17 kDa surface antigen precursor.
GN OMP OR RCL287.
OS Rickettsia conorii, and
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria;
OC Rickettsiales; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781, 783;
RN 11
RP SEQUENCE FROM N.A.
RC SPECIES-R.conorii, and R.rickettsii;
RX MEDLINE=89359171; PubMed=2768201;
RA Anderson B.E., Tzianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen
RL gene.";
RL J. Bacteriol. 171:5199-5201(1989).
RN 12
RP SEQUENCE FROM N.A.
RC SPECIES-R.conorii; STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samsom D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN 13
RP SEQUENCE FROM N.A.
RC SPECIES-R.rickettsii;
RX MEDLINE=87222152; PubMed=3108232;
RA Anderson B.E., Kennedy R.L., Carlone G.M., Tzianabos T., McDade J.E.,
RA Fu Z.Y., Bellini W.J.;
RT "Sequence analysis of the 17-kilodalton antigen gene from Rickettsia
RL rickettsii.";
RL J. Bacteriol. 169:2385-2390(1987).
RN 14
RP SEQUENCE OF 1-30 FROM N.A.
RC SPECIES-R.rickettsii;
RX MEDLINE=89080859; PubMed=3139629;
RA Anderson B.E., Baumstark B.R., Bellini W.J.;
RT "Expression of the gene encoding the 17-kilodalton antigen from
RL Rickettsia rickettsii: transcription and posttranslational
RL J. Bacteriol. 170:4493-4500(1988).
RN 15
RP -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
RP anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M28479; AAA26379.1;
DR EMBL, M28480; AAA26376.1;
DR EMBL, AE008675; AAL03825.1;
DR EMBL, M16486; AAA26381.1;
DR EMBL, J03371; -; NON_ANNOTATED_CDS.
DR PIR; A25972; A25972.
DR PIR; A31836; A31836.
DR PIR; A33971; A33971.
DR PIR; B33971; B33971.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KM Outer membrane: Lipoprotein; Antigen; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 159
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT CONFLICT 146 146 N->D (IN REF. 3).
FT CONFLICT 153 153 G->E (IN REF. 3).

```

[illegible]

```

Db      142 TTYGNACRQPDGQMRYN 159

RESULT 5
ID      17KD_RICAM STANDARD; PRT: 154 AA.
AC      P50928;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      17 kDa surface antigen precursor (Fragment).
OS      Rickettsia australis.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC      Rickettsiaceae; Rickettsiae; Rickettsia.
CX      NCBI_Taxid=787;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Baird R.W., Ross B., Dwyer B.;
RL      Submitted (Oct-1991) to the EMBL/GenBank/DDBJ databases.
CC      -I- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
        anchor (Probable).
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on ways
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
        or send an email to license@lsb-sib.ch).
-----
DR      EMBL; M74042; AAA26394.1; .
DR      PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW      Outer membrane; Lipoprotein; Antigen; Signal.
FT      SIGNAL 1 19 BY SIMILARITY.
FT      CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
FT      LIPID 20 20 N-ACTL DIGLYCERIDE (PROBABLE).
FT      NON_TER 154 154
SQ      SEQUENCE 154 AA; 15967 MW; E3AA833346FAC320 CRC64;

Query Match 33.0%; Score 276; DB 1; Length 154;
Best Local Similarity 38.1%; Pred. No. 4.9e-17;
Matches 59; Conservative 27; Mismatches 35; Indels 14; Gaps 4;

QY      9 SLLITSV---FLVGCAG--NFNRQEVGAATGAVVGAGOLFSGSGRVAMATGCAVLG 63
       I::II:::| | | | | : : : | | | | | : : | |
DB      5 SKIMITALAASLMQCNCPGCMNKGTGLTGAGAGALLGSFGKGKOIVGVCGALLG 64
       ::::| | | | | : : : | | | | | : : | |
QY      64 GLISGIQSOMDQDK---IKLNLSLEKVAKGYVTRMRNDPTGNSYSVEPVTRYQRNK 119
       ::::| | | | | : : : | | | | | : : | |
DB      65 AALGGQIAGCMDEDORRLAEILTSPRALETAPSSNVEMRPNDNGNYGVTPPKTVTRNSN 123
       ::::| | | | | : : : | | | | | : : | |
QY      120 QERRQVCREFQOKRMIAINGOKOEIYGTGRDPDR 154
       |||||:| | | | | : : : | | | | | : : | |
DB      124 ---GQYCRETYIVVIGKKQAIGNACROPDGO 154
       |||||:| | | | | : : : | | | | | : : | |

RESULT 6
ID      17KD_RICAM STANDARD; PRT: 154 AA.
AC      P50927;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      17 kDa surface antigen precursor (Fragment).
OS      Rickettsia amblyommali.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC      Rickettsiaceae; Rickettsiae; Rickettsia.
CX      NCBI_Taxid=33989;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      STRAIN-MO 85-1084;

```

RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
 CC Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U11013; AAB07704.1;
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KM Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19
 FT CHAIN 1 19
 FT LIPID 20 >154 BY SIMILARITY
 FT NON_TER 20 20 17 KDA SURFACE ANTIGEN.
 FT SEQUENCE 154 AA; 15879 MW; EAFBEAC29D943581 CRC64;
 Query Match 31.2%; Score 263; DB 1; Length 154;
 Best Local Similarity 36.8%; Pred. No. 6.4e-16;
 Matches 57; Conservative 27; Mismatches 57; Indels 14; Gaps 4;
 QY 9 SLLITISV---FLVGC--AQNFSROEVGAATGAVGVAGOLFGKSGSRVAMATGAVLG 63
 DB 5 SKIMIALAASMLQACNGPGMKNKOGTGLLGAGGALLSGSGFGKQOLVGVGVALLG 64
 QY 64 GLIGSKTIGSDMDOK-----IKLNSLEKVKAGVYTRMRNDPTGNSYSEVPVRYQRYNK 119
 DB 65 AVLGQIGAGMDQDRIELTSORALETAPSGSNVEMRNPDNGNIGYTPNKTY----- 119
 QY 120 QERRQYCRFEQOKAMIAQKQEIYGTACROPDGR 154
 DB 120 RNSTGYCREYQTQTVVIGGKQKATGNACLOPDQ 154

RESULT 7
 ID 17KD-RICPA STANDARD; PRT; 154 AA.
 AC P50930;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 17 kDa surface antigen precursor (Fragment).
 GN OMP.
 OS Rickettsia parkeri.
 CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiinae; Rickettsia.
 CC NCBI_TaxID=35792;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STROTHARD D.R., RALPH D.A., CLARK J.B., FUERST P.A., PRETZMAN C.;
 RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U17008; AAB82040.1;
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KM Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19
 FT CHAIN 1 19
 FT LIPID 20 >154 BY SIMILARITY
 FT NON_TER 20 20 17 KDA SURFACE ANTIGEN.
 FT SEQUENCE 154 AA; 15895 MW; OCF85AD5D96DFFEB CRC64;
 Query Match 31.2%; Score 261; DB 1; Length 154;
 Best Local Similarity 36.1%; Pred. No. 9.5e-16;
 Matches 56; Conservative 28; Mismatches 57; Indels 14; Gaps 4;
 QY 9 SLLITISV---FLVGC--AQNFSROEVGAATGAVGVAGOLFGKSGSRVAMATGAVLG 63
 DB 5 SKIMIALAASMLQACNGPGMKNKOGTGLLGAGGALLSGFGKQOLVGVGVALLG 64
 QY 64 GLIGSKTIGSDMDOK-----IKLNSLEKVKAGVYTRMRNDPTGNSYSEVPVRYQRYNK 119
 DB 65 AVLGQIGAGMDQDRIELTSORALETAPSGSNVEMRNPDNGNIGYTPNKTY----- 119
 QY 120 QERRQYCRFEQOKAMIAQKQEIYGTACROPDGR 154
 DB 120 RNSTGYCREYQTQTVVIGGKQKATGNACLOPDQ 154

RESULT 8
 ID 17KD-RICRH STANDARD; PRT; 154 AA.
 AC P50931;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 17 kDa surface antigen precursor (Fragment).
 GN OMP.
 OS Rickettsia rhipicephali.
 CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiinae; Rickettsia.
 CC NCBI_TaxID=33992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STROTHARD D.R., RALPH D.A., CLARK J.B., FUERST P.A., PRETZMAN C.;
 RL Submitted (AUG-1994) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U11020; AAB07706.1;
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KM Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19
 FT CHAIN 1 19
 FT LIPID 20 >154 BY SIMILARITY
 FT NON_TER 20 20 17 KDA SURFACE ANTIGEN.
 FT SEQUENCE 154 AA; 15895 MW; OCF85AD5D96DFFEB CRC64;
 Query Match 31.2%; Score 261; DB 1; Length 154;
 Best Local Similarity 36.1%; Pred. No. 9.5e-16;
 Matches 56; Conservative 28; Mismatches 57; Indels 14; Gaps 4;
 QY 9 SLLITISV---FLVGC--AQNFSROEVGAATGAVGVAGOLFGKSGSRVAMATGAVLG 63
 DB 5 SKIMIALAASMLQACNGPGMKNKOGTGLLGAGGALLSGFGKQOLVGVGVALLG 64
 QY 64 GLIGSKTIGSDMDOK-----IKLNSLEKVKAGVYTRMRNDPTGNSYSEVPVRYQRYNK 119
 DB 65 AVLGQIGAGMDQDRIELTSORALETAPSGSNVEMRNPDNGNIGYTPNKTY----- 119
 QY 120 QERRQYCRFEQOKAMIAQKQEIYGTACROPDGR 154
 DB 120 RNSTGYCREYQTQTVVIGGKQKATGNACLOPDQ 154

```

RESULT 9
17KD_RICMO STANDARD; PRT; 154 AA.
ID 17KD_RICMO
AC P50929;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia montana.
OC Bacteria: Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=33991;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OHIO 83-441;
RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U11017; AAB07705.1; -
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15881 MW; A09C53B8769E33DA CRC64;

Query Match 30.9%; Score 258; DB 1; Length 154;
Best Local Similarity 36.1%; Pred. No. 1.7e-15;
Matches 56; Conservative 28; Mismatches 57; Indels 14; Gaps 4;

OY 9 SLLIIV---FLVGC--AQNFSQEVGAATGAVGVAGQLFSGSGRYAMAIIGAVLG 63
DB 5 SKIMTIALASMLQACNGPGMNGGTGLGAGGALLSOFQGGQLVGVGVGALLG 64
OY 64 GLIGSKIGQSMDOQD---IKLNQSLKVKAGOVTRMRNDPTGNSYSVEPVRTYGRYK 119
DB 65 AVLAGGQIGAGMDEQDRLELTSSORALETAPSGSNVEMRNPDNGNYVTPNKTY----- 119

OY 120 QERRQYQREFQOKAMIAQKQEIYGTACROPDGR 154
DB 120 RNSGTGYCREYGTQVTVIGKQOKAYGNACLPDGO 154

RESULT 10
17KD_RICCA STANDARD; PRT; 80 AA.
ID 17KD_RICCA
AC P29697;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen (Fragment).
GN OMP.
OS Rickettsia canada.
OC Bacteria: Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=788;
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE=92108069; PubMed=1729713;
RX

```

```

RA Azad A.F., Sacchi J.B. Jr., Nelson W.M., Dasch G.A.,
RA Schmidtmann E.T., Carl M.;
RT "Genetic characterization and transovarial transmission of a
RT typhus-like rickettsia found in cat fleas.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:43-46(1992).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M82879; -; NOT_ANNOTATED_CDS.
DR Interpro: IPR000437; PROK_LIPOPROT.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
KW Outer membrane; Lipoprotein; Antigen.
FT NON_TER 1 1
FT NON_TER 80 80
SQ SEQUENCE 80 AA; 8372 MW; AD289A48EAB19E0E CRC64;

Query Match 18.8%; Score 157; DB 1; Length 80;
Best Local Similarity 40.3%; Pred. No. 4e-07;
Matches 31; Conservative 15; Mismatches 27; Indels 4; Gaps 1;

OY 43 GQLFGKSGRYAMAIIGAVLGIGLSKIGQSMDOQD---IKLNQSLKVKAGOVTRMR 98
DB 1 GSGFGKKGQGLIGVACALLGAILGNQIGAGMDEQDRLELTSSORALETTPSTSEWR 60

OY 99 NPDGTGNSYSVEPVRTYQ 115
DB 61 NPDGNGYGVTPSKTYK 77

RESULT 11
PCP_YEREN STANDARD; PRT; 155 AA.
ID PCP_YEREN
AC P31484;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane lipoprotein pcg precursor.
GN PCP OR PCPY.
OS Yersinia enterocolitica.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51872 / WA-C / SEROTYPE O:8;
RX MEDLINE=92121089; PubMed=1732192;
RA Baenleer A.J., Hanke K.;
RT "A lipoprotein of Yersinia enterocolitica facilitates ferrioxamine
RT uptake in Escherichia coli.";
RL J. Bacteriol. 174:1029-1035(1992).
CC -1- FUNCTION: FACILITATES FERRIOXAMINE UPTAKE.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -1- SIMILARITY: TO E. COLI AND S. TYPHIMURITUM SLI7B AND TO
CC H. INFLUENZAE PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X60448; CAA2977.1; -
DR

```

```

DR PIR: S23787; S23787.
DR HSSP: P00778; 1P04.
KW Outer membrane: Lipoprotein; 1.
FT SIGNAL 1 17
FT CHAIN 18 155
FT LIPID 18 155
SQ SEQUENCE 155 AA; 15362 MW; 8AD6B2132E849FA CRC64;

Query Match
Best Local Similarity 24.7%; Score 113.5; DB 1; Length 155;
Matches 33; Conservative 19; Mismatches 43; Indels 51; Gaps 4;

.OY 10 SLIISVFLVGCAGN-----FSRQE-----VGAT 34
Db 7 AVAIAVATLTCGANNNTLSGDVFSASQAKOVQTVYTGILSLRPVYTIQGDNNMGAIG 66
.OY 35 GAVVGVGAVGQLFGKSGRVMATIGAVLGLGSKIGSMDDQDKTKL----- 82
Db 67 GAVLGGFGLNTVGGGTGRSLATAGAVAGGAGGQVQAGMNRDGVLEVRKDDGTTILV 126
.OY 83 --NQLERVKAGQVTRWRNPDTCNSYSVDP 110
Db 127 VQKQGPTRFSVQ--RVMLASSGTVVSP 154

RESULT 12
SLIB_ECOLI
ID SLIB_ECOLI STANDARD: PRT: 155 AA.
AC P55741; P76183;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Outer membrane lipoprotein slib precursor.
GN SLIB OR B1641 OR 42655 OR ECS2350.
OS Escherichia coli, and
OC Escherichia coli O157:H7.
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-96133686; PubMed-8544813;
RA Ludwig A., Tengell C., Bauer S., Hubert A., Benz R., Mollenkopf H.-J.,
RA Goebel W.;
RT haemolytic and pore-forming protein in Escherichia coli.;
RT Mol. Gen. Genet. 249:474-486(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirtpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-K12;
RA MEDLINE-97251357; PubMed-9097039;
RA Albe H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kase H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Mori H., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Stampel G., Seki Y., Sivasubraman S.,
RA Tagami H., Takada J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RT DNA Res. 3:363-377(1996).
RN [4]

RP SEQUENCE FROM N.A.
RX STRAIN-O157:H7 / EDL933 / ATCC 700927;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirtpatrick H.A.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potomous K.,
RA Apodaca J., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RT Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN-O157:H7 / RIMD 0509552;
RA MEDLINE-21156231; PubMed-11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsuka E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RT DNA Res. 8:11-22(2001).
RN [6]
RP SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Potential).
CC - SIMILARITY: TO S. TYPHIMURIUM SLIB, H. INFLUENZAE PCP AND
CC Y. ENTEROCOLITICA PCP.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC use by non-profit institutions. There are no restrictions on its
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000259; AAC74713.1;
CC EMBL: D90807; BAA15402.1;
CC DR EMBL: AE005387; AAG56630.1;
CC DR EMBL: AP002558; BAB35773.1;
CC DR Ecogene; EG13409; SLIB
CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
CC KW Outer membrane: Lipoprotein; Signal; Complete proteome.
CC FT SIGNAL 1 17
CC FT CHAIN 18 155
CC FT LIPID 18 155
CC FT CONFLICT 98 18 N-ACYL DIGLYCERIDE.
CC SQ SEQUENCE 155 AA; 15602 MW; 543EB84069A5FA3 CRC64;

Query Match
Best Local Similarity 22.8%; Score 105.5; DB 1; Length 155;
Matches 33; Conservative 21; Mismatches 42; Indels 49; Gaps 4;

.OY 13 IISVFLVGCAGN-----FSRQE-----VGATGAV 37
Db 10 MVLGLVGVNNTLSGDVYTAISEAKQVQNTSYGTIVNAPVOIGGDSNVTGAIGAV 69
.OY 38 VGVVAGQLFGKSGRVMATIGAVLGLGSKIGSMDDQDKTKLNSLEKVKAGQVTRW 97
Db 70 LGFLENTVGGGTGRSLATAGAVAGGAGGQVQAGMNRDGVLEVRKDDGTTILV 126
.OY 98 RNDPTGNSYSVEPVYTRYNKQDR 122
Db 119 ---DDGNTINVVQKQGNTRFSRQK 140

RESULT 13
SLIB_SALTY
ID SLIB_SALTY STANDARD: PRT: 155 AA.
AC Q53549;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Outer membrane lipoprotein slib precursor.

```

GN SLVB OR STM1445 OR STY11677.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium;
RX MEDLINE=96133688; PubMed=8544813;
RA Ludwig A., Tengel C., Bauer S., Bubert A., Benz R., Mollenkopf H.-J.,
RA Goebel W.;
RT "SLVA, a regulatory protein from Salmonella typhimurium, induces a
RT haemolytic and pore-forming protein in Escherichia coli.";
RL Mol. Gen. Genet. 249:474-486(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Kettlewell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
CC -I- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Potential).
CC -I- SIMILARITY: TO E.COLI SLVB, H.INFLUENZAE PCP AND Y.ENTEROCOLITICA
CC PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: S80790; AAB35871.2; -;
CC DR EMBL: AE008262; AAL20367.1; -;
CC DR EMBL: AL627271; CAD01922.1; -;
CC DR StyGene: SG10573; slvb.
CC DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
CC KW Outer membrane; Lipoprotein; Signal; Complete proteome.
CC FT SIGNAL 1 17 POTENTIAL.
CC FT CHAIN 18 155 OUTER MEMBRANE LIPOPROTEIN SLVB.
CC FT LIPID 18 18 N-ACYL DIGLYCERIDE.
CC SQ SEQUENCE 155 AA; 15548 MW; 82PDDCCBBD5A7 CRC64;

Query Match 12.6%; Score 105.5; DB 1; Length 155;
Best Local Similarity 24.6%; Pred. No. 0.02;
Matches 34; Conservative 19; Mismatches 52; Indels 33; Gaps 3;

OY 4 GCGGSSSL-----ITISYFLVGCANFSRQEVGAATGAVGAGVACQ 44
DB 17 GCNNNDLSGVDYTASEAQOVNVTYGTIVNRPVQIQGGDDSNVIGAIAGVIGFLCN 76

OY 45 LFGKSGRVAMALGAVLGLGSKIGOSMDQODKIKLQNSLEKYKAGVTRMRNPDTCN 104
DB 77 TIGGTRGRSLATRAAGAVAGGAGGQVSAMNKIQVEL--EIRK-----DDGN 122

OY 105 SYSVEPRTYQRYNKOER 122
DB 123 TIMVVKQKGNTRFSAGQR 140

RESULT 14
PCP_HAEMIN
ID PCP_HAEMIN STANDARD; PRT; 155 AA.
AC P10325;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane lipoprotein PCP precursor (15 kDa lipoprotein) (PAL
DE cross-reacting lipoprotein).
GN PCP OR LPP OR H11579.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8811518; PubMed=2828309;
RA Delch R.A., Metcalf B.J., Finn C.W., Farley J.E., Green B.A.;
RT "Cloning of genes encoding a 15,000-dalton peptidoglycan-associated
RT outer membrane lipoprotein and an antigenically related 15,000-dalton
RT protein from Haemophilus influenzae.";
RL J. Bacteriol. 170:489-498(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ullrichback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -I- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -I- SIMILARITY: TO E.COLI AND S.TYPHIMURIUM SLVB AND TO
CC Y.ENTEROCOLITICA PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M18877; AAA24938.1; -;
CC DR EMBL: U32832; AAC23228.1; -;
CC DR PIR: B28543; B28543.
CC DR TIGR: H11579; -;
CC DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
CC KW Outer membrane; Lipoprotein; Signal; Complete proteome.
CC FT SIGNAL 1 18
CC FT CHAIN 19 155 OUTER MEMBRANE LIPOPROTEIN PCP.
CC FT LIPID 19 19 N-ACYL DIGLYCERIDE.
CC FT CONFLICT 135 143 CSLVAEVE -> VAGRRVRI (IN REF. 1).
CC SQ SEQUENCE 155 AA; 15425 MW; D7880327FCF0C985 CRC64;

Query Match
Best Local Similarity 12.2%; Score 102; DB 1; Length 155;
Matches 21; Conservative 9; Mismatches 23; Indels 0; Gaps 0;

QY 30 VGATGAVGAGVAGOLFQKSGSRVMAIGAVLGLSGKSGSMQDQDKIKI 82
DB 62 VGTGGGALGGTGGTGGGCGQATMAVGAIGALGSGKSGKSGKSGVNGAEL 114

RESULT 15
YCFJ_ECOLI

ID YCFJ_ECOLI STANDARD: PRT; 179 AA.
AC P37796; P375951;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
GN YCFJ OR B1110.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aida H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE OF 1-63 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=81236546; PubMed=6265208;
RA Young J.G., Rogers B.L., Campbell H.D., Jaworowski A., Shaw D.C.;
RT "Nucleotide sequence coding for the respiratory NADH dehydrogenase of
Escherichia coli using initiation codon.";
RL Eur. J. Biochem. 116:165-170(1981).
RN [4]
RP IDENTIFICATION
RX MEDLINE=95075659; PubMed=7984428;
RA Borodovsky M., Rudd K.E., Koonin E.V.;
RT "Intrinsic and extrinsic approaches for detecting genes in a
bacterial genome.";
RL Nucleic Acids Res. 22:4756-4767(1994).
RN [5]
RP SIMILARITY: TO RICKETTSIA 17 KDa SURFACE ANTIGEN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch.)

EMBL: AE000211; AAC74194.1; -
EMBL: D90746; BAA35925.1; -
EMBL: V00306; -; NOT_ANNOTATED_CDS.

DR Ecogene: EG12444; ycfJ.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 5 POTENTIAL.
SQ SEQUENCE 179 AA; 18920 MW; BASEBDBS6D45609 CRC64;

Query Match
Best Local Similarity 11.8%; Score 99; DB 1; Length 179;
Matches 26; Conservative 12; Mismatches 32; Indels 26; Gaps 1;

QY 31 GAATGAVGAGVAGOLFQKSGSRVMAIGAVLGLSGKSGSMQDQDKIKINQSLERVK 90
DB 73 GSVLGAAGAGVAGHOFGGGRGKDVATVYVAGLGGYAGNQLQSLQESD----- 120
QY 91 AGGVTRMRNPDTGNSYSVEPVRTYQRYNKRQRY 126
DB 121 -----TYTTTQGRCKTYVDKSEKMLGY 142

Search completed: October 27, 2002, 11:10:51
Job time: 7.15544 secs

C:Species: Rickettsia conorii
 C:Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 20-Aug-1999
 C:Accession: B33971
 R:Anderson, B.E.; Tzianabos, T.
 J. Bacteriol. 171, 5199-5201, 1989
 A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.
 A:Reference number: A33971; MUID:89359171
 A:Accession: B33971
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-159 <AND>
 A:Cross-references: GB:M28479; NID:9152463; PIDN:AAA2637.1; PID:9152464
 C:Superfamily: rickettsial common antigen

Query Match
 Best Local Similarity 35.0%; Score 293; DB 2; Length 159;
 Matches 61; Conservative 29; Mismatches 56; Indels 14; Gaps 4;
 QY 9 SLLITISV---FLVGC--AQNFSROEVGATGAVGVAGVAGQLFGKSGRVAMAGAVIG 63
 Db 5 SKIMTIALATSMLOACNGPGMKNKGCTGTLGAGAGALLGSGFGKGGQLVGVGALLG 63
 QY 64 GLIGSKIGQSMDOQDK----IKLNSLEKVKAGQVTRMRNPDPTGNSYSVEPVRTYQRYNK 119
 Db 65 AVLGQIGAGMDEQDRRLAELTSGRALETAPSGSNVEMRPNQNGYVTPNKTY----- 119
 QY 120 QERRQYCRFEOQKAMIAQOKQELTYGTACROPDGRMVOYIS 159
 Db 120 RNSTGQYCRETYTQTVVIGKQOKAYGNACROPDGMQOVYN 159

RESULT 3
 A33971
 Rickettsial common antigen precursor - Rickettsia rickettsii
 C:Species: Rickettsia rickettsii
 C:Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 20-Aug-1999
 C:Accession: A33971
 R:Anderson, B.E.; Tzianabos, T.
 J. Bacteriol. 171, 5199-5201, 1989
 A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.
 A:Reference number: A33971; MUID:89359171
 A:Accession: A33971
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-159 <AND>
 A:Cross-references: GB:M28480; NID:9152457; PIDN:AAA2637.1; PID:9152458
 C:Superfamily: rickettsial common antigen

Query Match
 Best Local Similarity 35.0%; Score 293; DB 2; Length 159;
 Matches 61; Conservative 29; Mismatches 56; Indels 14; Gaps 4;
 QY 9 SLLITISV---FLVGC--AQNFSROEVGATGAVGVAGVAGQLFGKSGRVAMAGAVIG 63
 Db 5 SKIMTIALATSMLOACNGPGMKNKGCTGTLGAGAGALLGSGFGKGGQLVGVGALLG 63
 QY 64 GLIGSKIGQSMDOQDK----IKLNSLEKVKAGQVTRMRNPDPTGNSYSVEPVRTYQRYNK 119
 Db 65 AVLGQIGAGMDEQDRRLAELTSGRALETAPSGSNVEMRPNQNGYVTPNKTY----- 119
 QY 120 QERRQYCRFEOQKAMIAQOKQELTYGTACROPDGRMVOYIS 159
 Db 120 RNSTGQYCRETYTQTVVIGKQOKAYGNACROPDGMQOVYN 159

RESULT 4
 G97860
 17K surface antigen precursor [Imported] - Rickettsia conorii (strain Malish 7)
 C:Species: Rickettsia conorii
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
 C:Accession: G97860
 R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
 science 293, 2033-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
 A:Reference number: A97700; MUID:21442074; PMID:11557893
 A:Accession: G97860
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-159 <KOR>
 A:Cross-references: GB:A6006914; PIDN:AAL03825.1; PID:915620425; GSPDB:GN00173
 C:Gene: omp
 C:Genetics:
 A:Superfamily: rickettsial common antigen

Query Match
 Best Local Similarity 35.0%; Score 293; DB 2; Length 159;
 Matches 61; Conservative 29; Mismatches 56; Indels 14; Gaps 4;
 QY 9 SLLITISV---FLVGC--AQNFSROEVGATGAVGVAGVAGQLFGKSGRVAMAGAVIG 63
 Db 5 SKIMTIALATSMLOACNGPGMKNKGCTGTLGAGAGALLGSGFGKGGQLVGVGALLG 63
 QY 64 GLIGSKIGQSMDOQDK----IKLNSLEKVKAGQVTRMRNPDPTGNSYSVEPVRTYQRYNK 119
 Db 65 AVLGQIGAGMDEQDRRLAELTSGRALETAPSGSNVEMRPNQNGYVTPNKTY----- 119
 QY 120 QERRQYCRFEOQKAMIAQOKQELTYGTACROPDGRMVOYIS 159
 Db 120 RNSTGQYCRETYTQTVVIGKQOKAYGNACROPDGMQOVYN 159

RESULT 5
 C33971
 Rickettsial common antigen precursor - Rickettsia typhi
 C:Species: Rickettsia typhi
 C:Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 20-Aug-1999
 C:Accession: C33971
 R:Anderson, B.E.; Tzianabos, T.
 J. Bacteriol. 171, 5199-5201, 1989
 A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.
 A:Reference number: A33971; MUID:89359171
 A:Accession: C33971
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-159 <AND>
 A:Cross-references: GB:M28481; NID:9152459; PIDN:AAA2637.1; PID:9152460
 C:Superfamily: rickettsial common antigen

Query Match
 Best Local Similarity 34.0%; Score 284.5; DB 2; Length 159;
 Matches 55; Conservative 25; Mismatches 49; Indels 9; Gaps --;
 QY 26 SROEVGATGAVGVAGVAGQLFGKSGRVAMAGAVIGGLIGSKIGQSMDOQDK----IK 81
 Db 27 NKQGTGTLGAGAGALLGSGFGKGGQLVGVGALLGSLDQDKLLELT 86
 QY 82 INQSLKVKAGQVTRMRNPDPTGNSYSVEPVRTYQRYNKQERRQYCRFEOQKAMIAQOKO 141
 Db 87 SORALESPAGSSNEMRPNQNGYVTPNKTY-----RNSTGQYCRETYTQTVVIGKQOK 141
 QY 142 EITYGTACROPDGRMVOYIS 159
 Db 142 TTYGNACROPDGMQOVYN 159

RESULT 6
 A25972
 17K antigen precursor - Rickettsia rickettsii
 C:Species: Rickettsia rickettsii
 C:Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 20-Aug-1999
 C:Accession: A25972
 R:Anderson, B.E.; Regnery, R.L.; Carlone, G.M.; Tzianabos, T.; Mcdade, J.E.; Fu, Z.Y.
 J. Bacteriol. 169, 2385-2390, 1987
 A:Title: Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia ricketts
 A:Reference number: A25972; MUID:87222152
 A:Accession: A25972

A:Molecule type: DNA
A:Residues: 1-159 <AND>
A:Cross-references: GB:M16486; NID:g152467; PIDN:AAA26381.1; PID:g152468
C:Superfamily: ticketsl common antigen

Query Match 34.0%; Score 284; DB 2; Length 159;
Best Local Similarity 37.5%; Pred. No. 7.7e-18;

Matches 60; Conservative 23; Mismatches 57; Indels 14; Gaps 4;

QY 9 SLLIIVY---FLVGC--AQNFSQEVGATGAVGAGVAGLFGKSGGRAVMAIGAVLG 63

Db 5 SKIMIIATSMLOACACPGPMNKQGTGLGAGGALLGSGFGKGGQGLVGVGALLG 64

QY 64 GLIGSKIGSGMDODK-----IKLMSLEKVKAGQVTRMRNPDNGNSVSEPVRTYGRYK 119

Db 65 AVLGQIGAGMDQDLAEITTSQRALETAPSGSNVEMRPNNGNNGYVTPNKTY----- 119

QY 120 QERRQYCRFPOKAMTAGQKEITYGTACRQPDGRQVTS 159

Db 120 RNSGQYCRFPTQVTVIGRQKQKAYGACRQPDGRQVTVN 159

RESULT 7
DB3169
conserved hypothetical protein PA3819 [Imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: DB3169
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mliloguchl, S.D.; Warren, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lm,

.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337

A:Accession: DB3169
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-182 <STO>

A:Cross-references: GB:AE004799; GB:AE004091; NID:g9949961; PIDN:AAG07206.1; GSPDB:GN001

A:Experimental source: strain PA01
C:Genetics:

A:Gene: PA3819

Query Match 15.4%; Score 128.5; DB 2; Length 182;
Best Local Similarity 34.8%; Pred. No. 0.0004;

Matches 32; Conservative 13; Mismatches 28; Indels 19; Gaps 2;

QY 28 QEVGAATGAVVAGVAGQLFGKSGRVAMAIGAVLGILGSKIGSMDODKI----- 80

Db 70 QIAGTATGAVVAGLGNQIGGCTKATATAGAVGGVAGNKNVQEGMQERDPTTETTRC 129

QY 81 KLNQSLKLV-----KAGQVTRMRNP 100

Db 130 STVHDSSEKVVGVYDKYMLDGKAGQIRMRNDP 161

RESULT 8
AI318
17k surface antigen precursor [Imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

C:Accession: AI318
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,

.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens

A:Reference number: AD35252; PMID:11756688
A:Accession: AI318
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-131 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL52516.1; PID:g17983328; GSPDB:GN00190
A:Experimental source: strain 16M

C:Genetics:
A:Gene: BME11335
A:Map position: 1

Query Match 15.0%; Score 125.5; DB 2; Length 131;
Best Local Similarity 31.7%; Pred. No. 0.00052;

Matches 39; Conservative 18; Mismatches 43; Indels 23; Gaps 7;

QY 47 GKSGGRAVMAIG-----AVLG-GLIGSKIGQ--SMDQDKIKLN-QSLEKVKAG 92

Db 14 GKSGG--FPSLGSSSQAPETNLASTLNGLENGNSASQLSADRRKALEAEVRALEYSBAG 71

QY 93 QVTRMRNPDNGNSVSEPVRTYGRYKQYCRFPOKAMTAGQKEITYGTACRQPDGRQVTS 152

Db 72 KSVLMSAGS--NAGDVTAAQYQ-----VGSQNCROYSHSFTTIGGQDQYVGTACRNP 124

QY 153 GRW 155

Db 125 GSW 127

RESULT 9
B83514

conserved hypothetical protein PA1053 [Imported] - Pseudomonas aeruginosa (strain PA0

C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: B83514
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mliloguchl, S.D.; Warren, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: B83514
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-154 <STO>

A:Cross-references: GB:AE004537; GB:AE004091; NID:g9946960; PIDN:AAG04442.1; GSPDB:GN

A:Experimental source: strain PA01
C:Genetics:

A:Gene: PA1053
C:Superfamily: PAL cross-reacting lipoprotein

Query Match 13.9%; Score 116; DB 2; Length 154;
Best Local Similarity 25.0%; Pred. No. 0.0042;

Matches 34; Conservative 21; Mismatches 41; Indels 40; Gaps 3;

QY 1 MNRCLGSSLIIVFLVGC-----AQNFSROE----- 29

Db 1 MKRSALIVASTAMALALGCGSSLTGDTYSREARIVQYRMGTIQALRPVKIEGTRP 60

QY 30 VGAATGAVVAGVAGQLFGKSGRVAMAIGAVLGILGSKIGSMDQDKIKL----- 82

Db 61 IGSIAGVAGVAGSAYGKGSYVAAIIGAVAGLGAATEBELTTRQGVETIVREDG 120

QY 83 --NQSLEKVKAGQVTR 96

Db 121 STRAYVOOVDDGQIFR 136

RESULT 10
S23787
outer membrane lipoprotein precursor - Yersinia enterocolitica

C:Species: Yersinia enterocolitica
C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 29-Sep-1999

C:Accession: S23787
R:Beumler, A.J.; Hantke, K.
J. Bacteriol. 174, 1029-1035, 1992

A:Title: A lipoprotein of Yersinia enterocolitica facilitates ferrioxamine uptake in

A:Reference number: S23787; MUID:92121089
A:Accession: S23787
A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-155 <BAE>
A:Cross-references: EMBL:X60448; NID:948577; PIDN:CAA42977.1; PID:948579
C:Superfamily: PAL cross-reacting lipoprotein

Query Match

Best Local Similarity 13.6%; Score 113.5; DB 2; Length 155;
Matches 37; Conservative 19; Mismatches 43; Indels 51; Gaps 4;

OY 10 SLIISVFLVCAQN-----FSROE-----
DB 7 AATAAATLTCGANNNTLSGPFSASAKQVOTVYTGILLSVPYPTIGCGDONNMKGHIG 34
OY 35 GAVVGVAGVQLFGKSGSRVAMAIAGAVLGLIGSKIGOSMDQDQIKL-----
DB 67 GAVLGGFLGNTVGGCTGRSLATTAAGAVAGMAGGCGVAGNMRTDGVOLVVRKDDGTTLIV 82
OY 83 --NOLSKVAKQVTRRRNPDGNSYSYEP 110
DB 127 VOKGPTRFVSQ--RVMGLASSGRTVYSP 154

RESULT 11

B82837

Conserved hypothetical protein Xf0178 [Imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence-revision 20-Aug-2000 #text-change 20-Aug-2000
A:Accession: B82837
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences, 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; PMID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: B82837
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-257 <SIM>

A:Cross-references: GB:AE003872; GB:AE003849; NID:99104975; PIDN:AAF82991.1; GSPDB:GN001

R:Experimental source: strain 9a5c
R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Bionet, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, H.

submitted to Genbank, June 2000
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Klieger, J.E.; Kuanene, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, E.C.; Miyaki, C.Y.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Contents: annotation
A:Gene: Xf0178

Query Match

Best Local Similarity 13.4%; Score 112; DB 2; Length 257;
Matches 20; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

OY 31 GAATGAVVGVAGVQLFGKSGSRVAMAIAGAVLGLIGSKI 70
DB 105 GTAICATLIGLGVGNQFGHGNKRKALTRAAGAVAGGFIGNEV 144

RESULT 12

AD2696

lipA protein [Imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence-revision 11-Jan-2002 #text-change 11-Jan-2002
A:Accession: AD2696
R:Wood, D.W.; Sebubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

erage, G.; Giller, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McClell
Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Ka
ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58
A:Reference number: AB2577; PMID:11743193
A:Accession: AD2696
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-142 <KUR>

A:Cross-references: GB:AE008688; PIDN:AAL41986.1; PID:g17739358; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)
C:Gene: lipA
A:Map position: circular chromosome

Query Match
Best Local Similarity 13.3%; Score 111.5; DB 2; Length 142;
Matches 41; Conservative 21; Mismatches 60; Indels 29; Gaps 9;

OY 10 SLIISVFLVCAQNFSGEATGAVVGVAGVQLFGKSGSRVAMAIAGAVLGLIGSK 69
DB 12 SLICVSM-LSAC-----TTGTRPAG--GSLFGR-SAQSTPFLANLGGIVG-K 56
OY 70 IGOSMDQDQIKL-----NOLSKVAKQVTRRRNPDGNSYSYEPVYTRYRNRKERRQ 124
DB 57 SGVELDGDQTKALEYKALFAPVGTPTGVDVAGVYANAP---YOVGN----- 107
OY 125 QYCNFQKAMIAQKQEIYGTACRQPDGRW 155
DB 108 QNCRQYSHLTVDGRTVRVGAACRNDGWSW 138

RESULT 13

D97478

lipA protein [Imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence-revision 30-Sep-2001 #text-change 11-Jan-2002
A:Accession: D97478
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194

A:Accession: D97478
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-125 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK6781.1; PID:g15155981; GSPDB:GN00169
A:Gene: AGR_C_1782
A:Map position: circular chromosome

Query Match

Best Local Similarity 13.3%; Score 111; DB 2; Length 125;
Matches 36; Conservative 18; Mismatches 61; Indels 16; Gaps 6;

OY 30 VCAATGAVVGVAGVQLFGKSGSRVAMAIAGAVLGLIGSKIGOSMDQDQIKL-----NOS 85
DB 2 LSACTTGTGTRPAGSIFGR-SAQSTPFLANLGGIVG-KSGVELDGDQTKALEYK 59
OY 86 LEKVKAGQVTRRRNPDGNSYSYEPVYTRYRNRKERRQYCNFQKAMIAQKQEIY 144
DB 60 LETAPVGTPTVITGDDVYKQVYANAP---YOVGN-----QNCROYSHTLTVDGRTVR 110
OY 145 GTACRQPDGRW 155
DB 111 GAACRNDGWSW 121

RESULT 14

S58234

lipA protein - Rhizobium leguminosarum
C:Species: Rhizobium leguminosarum

THIS PAGE BLANK (USPTO)

Tue Oct 29 08:30:54 2002

us-09-677-374-2.oligo.ra1

Db 491 OGAGAAAAAAGAGGCGTIRGGAGGCGYGGGSGSGRGGLGGGAGAAAAAGAGGAGG 550
QY 65 LIGSKIGQ 72
Db 551 LGGGAGG 558

RESULT 2

US-09-247-806-1
Sequence 1, Application US/09247806
Patent No. 6280747
GENERAL INFORMATION:
APPLICANT: PHILIPPE, Michel
APPLICANT: GARSON, Jean-Claude
APPLICANT: ARAUDEAU, Jean-Pierre
TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
FILE REFERENCE: 6388-0365-0
CURRENT APPLICATION NUMBER: US/09/247,806
EARLIER FILING DATE: 1999-02-11
EARLIER APPLICATION NUMBER: FR 98/01614
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO: 1
LENGTH: 651
TYPE: PR1
ORGANISM: Nephila clavipes
US-09-247-806-1

Query Match

Best Local Similarity 9.7%; Score 81.5; DB 4; Length 651;
Matches 25; Conservative 3; Mismatches 17; Indels 23; Gaps 2;

QY 28 OEVAAATGAVG---GVAGQ-----LFGGSGRVAMATGAVLGG 64
Db 491 OGAGAAAAAAGAGGCGTIRGGAGGCGYGGGSGSGRGGLGGGAGAAAAAGAGGAGG 550
QY 65 LIGSKIGQ 72
Db 551 LGGGAGG 558

RESULT 3

US-08-425-069-2
Sequence 2, Application US/08425069
Patent No. 5728810
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
TITLE OF INVENTION: HUMAN, MICHAEL B.
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESS: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5728810th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Ver. 2.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,069
FILING DATE: 19-APR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
1

NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-069-2

Query Match

Best Local Similarity 9.7%; Score 81.5; DB 1; Length 718;
Matches 25; Conservative 3; Mismatches 17; Indels 23; Gaps 2;

QY 28 OEVAAATGAVG---GVAGQ-----LFGGSGRVAMATGAVLGG 64
Db 491 OGAGAAAAAAGAGGCGTIRGGAGGCGYGGGSGSGRGGLGGGAGAAAAAGAGGAGG 550
QY 65 LIGSKIGQ 72
Db 551 LGGGAGG 558

RESULT 4

US-08-317-844B-2
Sequence 2, Application US/08317844B
Patent No. 5989894
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
TITLE OF INVENTION: HUMAN, MICHAEL B.
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESS: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5989894th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Ver. 2.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,844B
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 241-1300
TELEFAX: (703) 241-2848
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-317-844B-2

Query Match

9.7%; Score 81.5; DB 2; Length 718.

RESULT 6
US-07-803-633A-13

Patent No. 5369025
GENERAL INFORMATION:
APPLICANT: NAZERIAN, Keyvan

APPLICANT: YANAGIDA, No. 53690250ru
APPLICANT: OGAWA, Ryohei
APPLICANT: LI, YI
APPLICANT: LI, YI

NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:

ADDRESS: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 301 No. 5369025th Washington Street
CITY: Falls Church

STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTED ADDRESS FROM:

COMPTON READABLE FORM:
1 MEDIUM TYPE: Floppy disk
2
3 COMPUTER: IBM PC compatible
4
5 OPERATING SYSTEM: PC DOS/MS-DOS
6
7

```

1  CURRENT APPLICATION DATA:
2  SOFTWARE: PatentIn Release #1.00, Version #1.25
3  CURRENT APPLICATION DATA:
4  APPLICATION NUMBER: US-07/803,633A
5

```

FILING DATE: 19911210
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1644-103P

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 241-1300
TELEFAX: (703) 241-2848

TELEX: 248345
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
OFF

LENGTH: 603 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein

Query Match	93%	Score 78:	DB 1:	Length 865:
US-07-803-633A-13	"025CBA".13.			
	Process			

Best Local Similarity 25.48; Pred. No. 2.9;
Matches 36; Conservative 30; Mismatches 48; Indels 28; Gaps 7;

QY 17 FLVCAQNFESRQ - EVGCAATGAVVGVAGQLPFKQSGRV - - - - - AMAIGCAVLGSLIG 67
 688 FMNGLELFFNMGVQCAIGKVVVGAAGAIVSTISVSAPMSNPFALATGLIIAGIAGA 747

```
QY      68 SKIGSMDQDOKIKLN-----QSLEKVKACQVTRWRNPDGTGNSYSVEPVRTYORYNKQ 120
```

Db	748	AFL- AVRYVNKLKSNPMKALYPMITEVLKA- QATRELHGSEDD-----LERTSID	796
Oy	121	ERRQRYCRE-FOOKAMIAQOKO	141

```
Db      797  ERKLEAREMIKYMALVSAEER 818
```

[illegible]

Sequence ID: APPRIATION: 00/00000000
Patent NO. 6309861
GENERAL INFORMATION:
APPLICANT: Ambrosius, Dorothea

```

;
;
; APPLICANT: Rudolph, Rainer
; APPLICANT: Schaeffner, Joerg
; APPLICANT: Schwarz, Elisabeth
;

```

; TITLE OF INVENTION: Process for the production of naturally folded and secreted p
 ; FILE REFERENCE: Case 20379

CURRENT APPLICATION NUMBER: US/09/553,498
CURRENT FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: EP99107412.1
PRIOR FILING DATE: 1999-04-26
NUMBER OF SEQ ID NOS: 10
SEQ ID NO 8
LENGTH: 255
TYPE: PRN
ORGANISM: E. coli
US-09-553-498-8

Query Match
Best Local Similarity 9.3%; Score 77.5; DB 4; Length 255;
Matches 37; Conservative 28; Mismatches 81; Indels 31; Gaps 7;

Oy 6 GSSLIITSVFLVGCANFROEVAATGAVY---GGVAGOLFGKSGGVAMAGAV 61
Db 83 LQMTSLREEDPAMTYCARDYAG-WGCGTTVYSSGGGGGGGGSDIELTOSPAI 141
Oy 62 LGGLGSKTIGSDMDQDKIK-LN-----OSLEKRVAGQVTRMRNPDGNSI 106
Db 142 MSASPGKRVMTCSASSSVRYMMVQOKSGTSPKMTYDTSKLSSGVPAFSGSGSTY 201
Oy 107 SVEFVRYQRYNKOEROOYCEFOOKMI--ACQKOETGTACRQPDGRWVISTE 161
Db 202 SL-----TISMEADATYTCQWMSNPLFGACTKLEKRAAEO-----KLISEE 249

RESULT 8
US-09-067-351-2
Sequence 2, Application US/09067351
Patent No. 5994081
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Baughn, Mariah
TITLE OF INVENTION: HUMAN KERATINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/067,351
FILING DATE: Herewith
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0511 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 551 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: KERANOT02
CLONE: 2029060
US-09-067-351-2

Query Match
Best Local Similarity 9.2%; Score 77; DB 2; Length 551;
Matches 28; Conservative 28; Mismatches 53; Indels 16; Gaps 2;

Oy 8 GSSLIITSVFLVGCANFROEVAATGAVGAGVAGOLFSGKSGGVAMAGAVLGLIG 67
Db 54 GASFGSRSLYMLGAKRVSLNCGSSCRSGFGGRASNGFVNSG---FGYGGVGSGFGSG 110
Oy 68 SK-----IGSDMDQDKIKLNOSLEKRVAGQVTRMRNPDGNSYVEFVRY 114
Db 111 PSFVPCPGCIQEVTVNOSLPLHLQIDPTIQRVAREERIKTLNNFTSFIDVRL 170
Oy 115 QRYK 119
Db 171 EQOK 175

RESULT 9
US-09-360-490-2
Sequence 2, Application US/09360490
Patent No. 6221843
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Baughn, Mariah
TITLE OF INVENTION: HUMAN KERATINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/360,490
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/067,351
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0511 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 551 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: KERANOT02
CLONE: 2029060
US-09-360-490-2

Query Match
Best Local Similarity 22.4%; Score 77; DB 4; Length 551;
Matches 28; Conservative 28; Mismatches 53; Indels 16; Gaps 2;

Oy 8 GSSLIITSVFLVGCANFROEVAATGAVGAGVAGOLFSGKSGGVAMAGAVLGLIG 67
Db 54 GASFGSRSLYMLGAKRVSLNCGSSCRSGFGGRASNGFVNSG---FGYGGVGSGFGSG 110

Qy 68 SK-----IGSQMDQDKIKLNQSLKRVKAGVTRMNPDTGNSYSEPVRTY 114
Db 111 PSPFVCPGIGQEVTVNQSILTPHLQIDPIQIVRAERBQKITLNNKTFISIDKVRFL 170
Qy 115 QRYNK 119
Db 171 EQQNK 175

RESULT 10
US-08-374-077C-2
Sequence 2, Application US/08374077C
Patent No. 6027912
GENERAL INFORMATION:
APPLICANT: Hall, Linda M.
APPLICANT: Ren, Dejian
APPLICANT: Zheng, Wei
APPLICANT: Dubald, Manuel Marcel Paul
TITLE OF INVENTION: Genes Encoding an Invertebrate Alpha
TITLE OF INVENTION: Calcium Channel Subunit
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/374, 077C
FILING DATE: 19-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm M.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 022650-264
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2516 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-374-077C-2

Query Match 9.2%; Score 76.5; DB 3; Length 2516;
Best Local Similarity 24.8%; Pred. No. 17;
Matches 32; Conservative 18; Mismatches 66; Indels 13; Gaps 4;

Qy 30 VGAATGAVGVAGQLFGKSGRYAMAIG--AVLGLISKIGQS---MDQDKIKLN 83
Db 2389 IGSSNGSIFGSGAGLGAGSGGVG-GLGSSSIRNAFGSGSPSSLPQHGYSGTLN 2447
Qy 84 -----QSLKRVKAGVTRMNPDTGNSYSEPVRTYQRINKORROOYCREFOQKAMIA 137
Db 2448 SPPIDNRRLRRVATVTTNNNNKSQVSNSSSLNVRANNSQNMSPGQPVQOQSPLR 2507
Qy 138 GQKOEIYGT 146
Db 2508 GQGNQTYSS 2516

RESULT 11
US-08-895-590-2
Sequence 2, Application US/08895590
Patent No. 6207410

GENERAL INFORMATION:
APPLICANT: Hall, Linda M.
APPLICANT: Ren, Dejian
APPLICANT: Zheng, Wei
APPLICANT: Dubald, Manuel Marcel Paul
TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895, 590
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/374, 888
FILING DATE: 19-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm M.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 022650-263
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2516 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-895-590-2

Query Match 9.2%; Score 76.5; DB 4; Length 2516;
Best Local Similarity 24.8%; Pred. No. 17;
Matches 32; Conservative 18; Mismatches 66; Indels 13; Gaps 4;

Qy 30 VGAATGAVGVAGQLFGKSGRYAMAIG--AVLGLISKIGQS---MDQDKIKLN 83
Db 2389 IGSSNGSIFGSGAGLGAGSGGVG-GLGSSSIRNAFGSGSPSSLPQHGYSGTLN 2447
Qy 84 -----QSLKRVKAGVTRMNPDTGNSYSEPVRTYQRINKORROOYCREFOQKAMIA 137
Db 2448 SPPIDNRRLRRVATVTTNNNNKSQVSNSSSLNVRANNSQNMSPGQPVQOQSPLR 2507
Qy 138 GQKOEIYGT 146
Db 2508 GQGNQTYSS 2516

RESULT 12
US-07-910-760-12
Sequence 12, Application US/07910760
Patent No. 5683864
GENERAL INFORMATION:
APPLICANT: Houghton, Michael
APPLICANT: Choo, Qui-Lim
APPLICANT: Kuo, George
TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
TITLE OF INVENTION: Antigens for use in Immunoassays for Anti-HCV Antibodies
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: P.O. Box 8097 (Int. Prop. R-440)
CITY: Emeryville

STATE: CA
 COUNTRY: U.S.A.
 ZIP: 94662-8097
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/910,760
 FILING DATE: 07-JUL-1992
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Blackburn Esq., Robert P.
 REGISTRATION NUMBER: 30,447
 REFERENCE/DOCKET NUMBER: 0101.002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (510) 601-2702
 TELEFAX: (510) 655-3542
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1021 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-910-760-12

Query Match
 Best Local Similarity 9.1%; Score 76; DB 1; Length 1021;
 Matches 42; Conservative 24; Mismatches 63; Indels 72; Gaps 9;

QY 31 GAATGAVVGVAGQLFGK-GSGRVMAI-----GGAVLGLGLGSKI--GQSMDOQDKIKIN 83
 DB 792 GAATAFVAGAGAGAAIGSVGLKVLIDILAGYAGAGALVAFKIMGSEVPSTEDLVNLL 851
 QY 84 QSL-----EKVAKQ-VTRWRN-----PDGNSYSVEPVRTYQRYNKQ 120
 DB 852 PATLSPGALVGVCAAILRRHVGPBGAVQWMNRLIAFASRGNHVSPGNSSTNKPQK 911
 QY 121 ERRO-----QYCEPQOKAMIAQCKOETIYTAC 148
 DB 912 NKRTNRRPDPVKPPGGQIVGVYLLPRGRPLGVRAIRKTSERSOPRGRQPI--PKA 969
 QY 149 ROPDGR-----WQYISTE 161
 DB 970 RRPEGRVWAGQGYPMPLYGNE 990

RESULT 13
 US-08-440-519-12
 Sequence 12, Application US/08440519
 Patent No. 5712087
 GENERAL INFORMATION:
 APPLICANT: Houghton, Michael
 APPLICANT: Choo, Qui-Lim
 TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Chiron Corporation
 STREET: P.O. Box 8097 (Int. Prop. R-440)
 CITY: Emeryville
 STATE: CA
 COUNTRY: U.S.A.
 ZIP: 94662-8097
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/440,519

FILING DATE: 12-MAY-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/910,760
 FILING DATE: 07-JUL-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Blackburn Esq., Robert P.
 REGISTRATION NUMBER: 30,447
 REFERENCE/DOCKET NUMBER: 0101.002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (510) 601-2702
 TELEFAX: (510) 655-3542
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1021 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-440-519-12

Query Match
 Best Local Similarity 9.1%; Score 76; DB 1; Length 1021;
 Matches 42; Conservative 24; Mismatches 63; Indels 72; Gaps 9;

QY 31 GAATGAVVGVAGQLFGK-GSGRVMAI-----GGAVLGLGLGSKI--GQSMDOQDKIKIN 83
 DB 792 GAATAFVAGAGAGAAIGSVGLKVLIDILAGYAGAGALVAFKIMGSEVPSTEDLVNLL 851
 QY 84 QSL-----EKVAKQ-VTRWRN-----PDGNSYSVEPVRTYQRYNKQ 120
 DB 852 PATLSPGALVGVCAAILRRHVGPBGAVQWMNRLIAFASRGNHVSPGNSSTNKPQK 911
 QY 121 ERRO-----QYCEPQOKAMIAQCKOETIYTAC 148
 DB 912 NKRTNRRPDPVKPPGGQIVGVYLLPRGRPLGVRAIRKTSERSOPRGRQPI--PKA 969
 QY 149 ROPDGR-----WQYISTE 161
 DB 970 RRPEGRVWAGQGYPMPLYGNE 990

RESULT 14
 US-08-440-549-12
 Sequence 12, Application US/08440549
 Patent No. 6312889
 GENERAL INFORMATION:
 APPLICANT: Houghton, Michael
 APPLICANT: Choo, Qui-Lim
 TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Chiron Corporation
 STREET: P.O. Box 8097 (Int. Prop. R-440)
 CITY: Emeryville
 STATE: CA
 COUNTRY: U.S.A.
 ZIP: 94662-8097
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/440,549
 FILING DATE: 12-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/910,760
 FILING DATE: 07-JUL-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Blackburn Esq., Robert P.

```

1      RESULT 15
2      US-08/190-199A-65
3      ; Sequence 65 Application US/08190199A
4      ; Patent No. 5830663
5      ; GENERAL INFORMATION:
6      ; APPLICANT: EMBLETON, Michael J.
7      ; APPLICANT: GOROKHOV, Guy
8      ; APPLICANT: JONES, Peter T.
9      ; APPLICANT: WINTER, Gregory P.
10     ; TITLE OF INVENTION: TREATMENT OF CELL POPULATIONS
11     ; NUMBER OF SEQUENCES: 70
12     ; CORRESPONDENCE ADDRESS:
13     ; ADDRESSEE: PILLSBURY MADISON & SUTRO, L.L.P.
14     ; STREET: 1100 New York Avenue, N.W.
15     ; CITY: Washington
16     ; STATE: D.C.
17     ; COUNTRY: U.S.A.
18     ; ZIP: 20005-3918
19     ; COMPUTER READABLE FORM:
20     ; MEDIUM TYPE: Floppy disk
21     ; COMPUTER: IBM PC compatible
22     ; OPERATING SYSTEM: PC-DOS/MS-DOS
23     ; SOFTWARE: Microsoft Word
24     ; CURRENT APPLICATION DATA:
25     ; APPLICATION NUMBER: US/08/190,199A
26     ; FILING DATE: 13-JUL-1994
27     ; CLASSIFICATION: 435
28     ; PRIOR APPLICATION DATA:
29     ; APPLICATION NUMBER: PCT/GB92/01483
30     ; FILING DATE: 10-AUG-1992
31     ; PRIOR APPLICATION DATA:
32     ; APPLICATION NUMBER: GB 9212419.7
33     ; FILING DATE: 11-JUN-1992
34     ; PRIOR APPLICATION DATA:
35     ; APPLICATION NUMBER: GB 9117352.6
36     ; FILING DATE: 10-AUG-1991
37     ; INFORMATION FOR SEQ. ID NO. 65:
38     ; SEQUENCE CHARACTERISTICS:
39     ; LENGTH: 236 amino acids

```

Search completed: October 27, 2002, 11:12:12
Job time : 9.39378 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:58:51 ; Search time 21.544 Seconds
(without alignments)
835.217 Million cell updates/sec

Title: US-09-677-374-2

Perfect score: 836
Sequence: 1 MNRGLQSGSLITISVFLVG.....ITGTACRQPDGRHMYISTEK 162

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq-032802:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	836	100.0	162	22	AAG78025	Piscirickettsia sa
2	836	100.0	162	22	AA881126	Ospa antigen amino
3	815	97.5	161	22	AA881127	Optimised Ospa pro
4	815	97.5	256	22	AA881128	CI7E2 Ospa constru
5	112	13.4	20	22	AA881130	Ospa B-cell epitop
6	108	12.9	224	22	AA820105	Moraxella catarrha
7	105.5	12.6	223	20	AA934487	Porphyromonas ging
8	105.5	12.6	230	20	AA934362	Porphyromonas ging
9	102	12.2	154	11	AA805799	PBOHP-2 gene prod.
10	98	11.7	309	22	AA815906	Novel human diagno
11	91	10.9	528	22	AA882611	Spider recombinant

12	86.5	10.3	2309	22	AB866232	Drosophila melanog
13	82	9.8	666	22	AB858019	Drosophila melanog
14	81.5	9.7	651	20	AA400697	Spider silk protei
15	81.5	9.7	718	12	AA814308	N.clavipes draglin
16	81.5	9.7	718	19	AA853346	Nephila clavipes s
17	81.5	9.7	718	21	AA859070	N. clavipes spider
18	80.5	9.6	102	22	AA841943	Human polyptide
19	80.5	9.6	302	22	AA840157	Human polyptide
20	80.5	9.6	542	22	AB865790	Drosophila melanog
21	80.5	9.6	542	22	AB865791	Drosophila melanog
22	80.5	9.6	542	22	AB870501	Drosophila melanog
23	80	9.6	116	19	AA811028	H. pylori ORF O1cp
24	80	9.6	2017	22	AB806301	Novel human diagno
25	80	9.6	2599	21	AA875098	Neisseria meningit
26	79	9.4	618	21	AA856803	Human prostate can
27	78	9.3	646	18	AA827178	Nephila clavipes s
28	78	9.3	865	14	AA830169	Marek's Disease VI
29	77.5	9.3	255	21	AA811398	E. coli expression
30	77.5	9.3	255	22	AA874199	PeLB-scFvOxazolom
31	77.5	9.3	255	22	AA870769	Expression plasmid
32	77.5	9.3	255	22	AA872020	E. carotovora PeLB
33	77.5	9.3	900	22	AB828648	Novel human diagno
34	77	9.2	147	22	AB869847	Drosophila melanog
35	77	9.2	429	10	AA890064	Antigenic protein
36	77	9.2	551	21	AA852398	Human Keratin KERT
37	77	9.2	1251	22	AB861254	Drosophila melanog
38	76.5	9.2	514	22	AA836520	Pseudomonas aerugi
39	76.5	9.2	2516	17	AA801875	Neuronal Invertebr
40	76.5	9.2	2516	17	AA801884	Invertebrate calci
41	76.5	9.2	2516	22	AB861076	Drosophila melanog
42	76	9.1	300	11	AA805878	Mercozite surface a
43	76	9.1	324	22	AA876751	Corynebacterium gl
44	76	9.1	346	22	AA892057	C glutamicum prote
45	76	9.1	346	22	AA876750	Corynebacterium gl

ALIGNMENTS

RESULT 1
ID AAG78025 standard; Protein; 162 AA.
XX AAG78025;
XX AC AAG78025;
XX DT 15-JAN-2002 (first entry)
XX DE Piscirickettsia salmonis polypeptide p10.6.
XX AC Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
XX KW septicemia; SRS; surface antigen; vaccine; antibacterial; fish;
XX ATCC VR-1361.
XX OS Piscirickettsia salmonis.
XX PN WO200168865-A2.
XX PD 20-SEP-2001.
XX PF 12-MAR-2001; 2001WO-GH01055.
XX PR 11-MAR-2000; 2000GB-0005838.
XX PR 01-JUL-2000; 2000GB-0016080.
XX PR 01-JUL-2000; 2000GB-0016082.
XX PR 29-JUL-2000; 2000GB-0018599.
XX PA (AQUA-) AQUA HEALTH EURO LTD.
XX PI Simard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;
XX PI Burzio L;
XX DR WPI: 2001-639050/73.
XX DR N-PSDB; AA879040.

```

XX New nucleic acids encoding an amino acid sequence homologous to the
PT surface antigen present on Piscirickettsia salmonis are useful to
PS protect fish against piscirickettsiosis
XX
XX Claim 6; Fig 5; 25pp; English.
XX
CC The invention relates to nucleic acid sequences and the encoded protein
CC of a least part of the surface antigen present on Piscirickettsia
CC salmonis for production of a vaccine with antibacterial activity to
CC protect fish against P. salmonis which causes piscirickettsiosis, also
CC known as salmonid rickettsial septicaemia.
XX
SO Sequence 162 AA:

Query Match
Best Local Similarity 100.0%; Score 836; DB 22; Length 162;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGCLGSSLIITISVFLVGCANFSEROEVAATGAVVGAAGLFGKSGRYAMATGCA 60
DB 1 MNRGCLGSSLIITISVFLVGCANFSEROEVAATGAVVGAAGLFGKSGRYAMATGCA 60
QY 61 VLGLIGSKIGGSMDDQDKIKINOSLEKVKAGVTRMRNPDGTGNSYSVEPVRTYORYNKO 120
DB 61 VLGLIGSKIGGSMDDQDKIKINOSLEKVKAGVTRMRNPDGTGNSYSVEPVRTYORYNKO 120
QY 121 ERROQYCFREFOOKAMIAQOKOETIGTACROPDGRMVOISTEK 162
DB 121 ERROQYCFREFOOKAMIAQOKOETIGTACROPDGRMVOISTEK 162

RESULT 2
AAB81126
ID AAB81126 standard; Protein; 162 AA.
XX
AC AAB81126;
XX
DT 11-JUL-2001 (first entry)
XX
DE OSPA antigen amino acid sequence.
XX
KW Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OSPA; salmonid rickettsial septicaemia; rickettsial disease;
XX
OS Piscirickettsia salmonis.
XX
FH Key
FT Location/Qualifiers
FT 110..129
FT /label=B_cell_epitope
XX
CA2281913-A1.
XX
PD 17-MAR-2001.
XX
PF 17-SEP-1999; 99CA-2281913.
XX
PR 17-SEP-1999; 99CA-2281913.
XX
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZV/) KUZUK M A.
XX
PI Kay WW, Burian J, Kuzuk MA;
XX
DR WPI: 2001-316844/34.
XX
DR N-PSDB: AAF86246.
XX
PT Method for protecting poikilothermic fish against salmonid rickettsial
PT septicemia and other rickettsial diseases comprises administering a
PT vaccine containing the OSPA protein of Piscirickettsia salmonis
XX

```

```

PS Example 2; Fig 2b; 35pp; English.
XX
XX This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed OSPA, or an immunogenic fragment of a
CC P. salmonis in the form of a vaccine. The method is used for protecting
CC particularly poikilothermic fish, against the bacterial pathogen
CC rickettsial septicaemia (SRS) and other rickettsial diseases.
CC Sequence represents P. salmonis OSPA protein. An OSPA protein with an
CC N-terminal fusion partner is used in a vaccine to create an anti-OSPA
XX antibody response.
XX
SO Sequence 162 AA:

Query Match
Best Local Similarity 100.0%; Score 836; DB 22; Length 162;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGCLGSSLIITISVFLVGCANFSEROEVAATGAVVGAAGLFGKSGRYAMATGCA 60
DB 1 MNRGCLGSSLIITISVFLVGCANFSEROEVAATGAVVGAAGLFGKSGRYAMATGCA 60
QY 61 VLGLIGSKIGGSMDDQDKIKINOSLEKVKAGVTRMRNPDGTGNSYSVEPVRTYORYNKO 120
DB 61 VLGLIGSKIGGSMDDQDKIKINOSLEKVKAGVTRMRNPDGTGNSYSVEPVRTYORYNKO 120
QY 121 ERROQYCFREFOOKAMIAQOKOETIGTACROPDGRMVOISTEK 162
DB 121 ERROQYCFREFOOKAMIAQOKOETIGTACROPDGRMVOISTEK 162

RESULT 3
AAB81127
ID AAB81127 standard; Protein; 161 AA.
XX
AC AAB81127;
XX
DT 11-JUL-2001 (first entry)
XX
DE Optimised OSPA protein 17E2 amino acid sequence.
XX
KW Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OSPA; salmonid rickettsial septicaemia; rickettsial disease;
XX
OS Piscirickettsia salmonis.
XX
FH Key
FT Location/Qualifiers
FT 109..128
FT /label=B_cell_epitope
XX
CA2281913-A1.
XX
PD 17-MAR-2001.
XX
PF 17-SEP-1999; 99CA-2281913.
XX
PR 17-SEP-1999; 99CA-2281913.
XX
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZV/) KUZUK M A.
XX
PI Kay WW, Burian J, Kuzuk MA;
XX
DR WPI: 2001-316844/34.
XX
DR N-PSDB: AAF86247.
XX
PT Method for protecting poikilothermic fish against salmonid rickettsial
PT septicemia and other rickettsial diseases comprises administering a
PT vaccine containing the OSPA protein of Piscirickettsia salmonis
XX

```


PT vaccine containing the Ospa protein of *Piscirickettsia salmonis* -
XX
PS Disclosure; Fig 5; 35pp; English.
CC This invention relates to a method for the protection against infection
CC of a polkilothermic fish by the bacterial pathogen, *Piscirickettsia*
CC *salmonis*. The method comprises administering an immunogenic amount of a
CC *P. salmonis* specific antigen termed Ospa, or an immunogenic fragment of a
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkilothermic fish, against the bacterial pathogen
CC *P. salmonis*. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents optimised *P. salmonis* Ospa protein 17E2. The DNA
CC encoding Ospa 17E2 (AA66247) has been optimised for expression in
CC *Escherichia coli*. An Ospa protein with an N-terminal fusion partner is
CC used in a vaccine to create an anti-Ospa antibody response.
XX
SQ Sequence 161 AA;
Query Match 97.5%; Score 815; DB 22; Length 161;
Best Local Similarity 98.8%; Pred. No. 5.1e-78;
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 3 RGCLGSSLLIIISVFLGCAQNFSEVGAATGAVGVAGOLFSGSGRYAMAIGAVL 62
DB 2 RGCLGSSLLIIISVFLGCAQNFSEVGAATGAVGVAGOLFSGSGRYAMAIGAVL 61
OY 63 GGLIGSKIGSMQDQDKIKLNQSLSEKVKAGVTRWRNPDTGNSYSVEPVRTYQRYNKOER 122
DB 62 GGLIGSKIGSMQDQDKIKLNQSLSEKVKAGVTRWRNPDTGNSYSVEPVRTYQRYNKOER 121
OY 123 RQCYCREFOOKAMTAGOKOETGTACROPDGRWVISTEK 162
DB 122 RQCYCREFOOKAMTAGOKOETGTACROPDGRWVISTEK 161
RESULT 4
AAB81128
ID AAB81128 standard; Protein: 256 AA.
AC AAB81128;
XX
DT 11-JUL-2001 (first entry)
XX
DE C17E2 Ospa construct with N-terminal fusion partner.
XX
KW Polkilothermic fish; *Piscirickettsia salmonis*; rickettsial pathogen;
KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
XX
XX SRS; 17E2; fusion construct.
OS *Piscirickettsia salmonis*.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Region 1..95 /label= Undefined_N-terminal_fusion_partner
FT 96..256 /label= C17E2_Ospa
FT Region /note= "Product of Ospa gene optimised for expression in
FT *Escherichia coli*"
XX
XX CA2281913-A1.
PN 17-MAR-2001.
PD 17-SEP-1999; 99CA-2281913.
PF 17-SEP-1999; 99CA-2281913.
PR 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
XX (BURI/) BURIAN J.
XX (KUZV/) KUZYK M A.
XX

PI Kay WM, Burian J, Kuzyk MA;
XX
XX WPI; 2001-316844/34.
DR N-PDB; AAF86248.
XX
XX Method for protecting polkilothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of *Piscirickettsia salmonis* -
XX
XX Example 4; Fig 5; 35pp; English.
PS
XX
CC This invention relates to a method for the protection against infection
CC of a polkilothermic fish by the bacterial pathogen, *Piscirickettsia*
CC *salmonis*. The method comprises administering an immunogenic amount of a
CC *P. salmonis* specific antigen termed Ospa, or an immunogenic fragment of a
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkilothermic fish, against the bacterial pathogen
CC *P. salmonis*. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents the amino acid sequence of C17E2, a *P. salmonis* Ospa
CC construct optimised for expression in *Escherichia coli*, fused to an
CC undefined N-terminal fusion partner. The fusion protein is used in a
CC vaccine to create an anti-Ospa antibody response.
XX
SQ Sequence 256 AA;
Query Match 97.5%; Score 815; DB 22; Length 256;
Best Local Similarity 98.8%; Pred. No. 9.3e-78;
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 3 RGCLGSSLLIIISVFLGCAQNFSEVGAATGAVGVAGOLFSGSGRYAMAIGAVL 62
DB 97 RGCLGSSLLIIISVFLGCAQNFSEVGAATGAVGVAGOLFSGSGRYAMAIGAVL 156
OY 63 GGLIGSKIGSMQDQDKIKLNQSLSEKVKAGVTRWRNPDTGNSYSVEPVRTYQRYNKOER 122
DB 157 GGLIGSKIGSMQDQDKIKLNQSLSEKVKAGVTRWRNPDTGNSYSVEPVRTYQRYNKOER 216
OY 123 RQCYCREFOOKAMTAGOKOETGTACROPDGRWVISTEK 162
DB 217 RQCYCREFOOKAMTAGOKOETGTACROPDGRWVISTEK 256
RESULT 5
AAB81130
ID AAB81130 standard; Peptide: 20 AA.
AC AAB81130;
XX
DT 11-JUL-2001 (first entry)
XX
XX Ospa B-cell epitope peptide #2.
XX
KW Polkilothermic fish; *Piscirickettsia salmonis*; rickettsial pathogen;
KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
XX
XX SRS; antibody.
OS *Piscirickettsia salmonis*.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Region 1..95 /label= Undefined_N-terminal_fusion_partner
FT 96..256 /label= C17E2_Ospa
FT Region /note= "Product of Ospa gene optimised for expression in
FT *Escherichia coli*"
XX
XX CA2281913-A1.
PN 17-MAR-2001.
PD 17-SEP-1999; 99CA-2281913.
PF 17-SEP-1999; 99CA-2281913.
PR 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
XX (BURI/) BURIAN J.
XX (KUZV/) KUZYK M A.
XX
XX Kay WM, Burian J, Kuzyk MA;
XX

DR WPI: 2001-316844/34.

XX Method for protecting poikilothermic fish against salmonid rickettsial
PT septicemia and other rickettsial diseases comprises administering a
XX vaccine containing the OSPA protein of *Piscirickettsia salmonis*
PS Example 2; Page 17; 35pp; English.

XX This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*
CC salmonis. The method comprises administering an immunogenic amount of a
CC *P. salmonis* specific antigen termed OSPA, or an immunogenic fragment of a
CC OSPA in the form of a vaccine. The method is used for protecting animals,
CC particularly poikilothermic fish, against the bacterial pathogen
CC *P. salmonis*. The method is also useful for protecting against salmonid
CC rickettsial septicemia (SR) and other rickettsial diseases. The present
CC sequence represents an immunogenic epitope of the *P. salmonis* OSPA
CC protein. The peptide is used to raise rabbit anti-OSPA antibodies.
XX Sequence 20 AA;

Query Match 13.4%; Score 112; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 4,9e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 110 PVRTQRYNKKQRRQOYCRE 129
Db 1 PVRTQRYNKKQRRQOYCRE 20

RESULT 6

AAB20105 standard; Protein; 224 AA.

XX AAB20105;

XX 23-APR-2001 (first entry)

XX Moraxella catarrhalis BASB13 protein.

XX BASB13; infection; otitis media; pneumonia; therapy; diagnosis;
KW antibacterial; antimicrobial.

XX Moraxella catarrhalis.

XX WO200100836-A1.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000MO-EP05851.

XX 25-JUN-1999; 99GB-0015044.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Thonnard J;

XX WPI: 2001-112458/12.

XX N-PSDB; AAF30043.

XX New BASB13 polypeptide isolated from Moraxella catarrhalis bacterium,
PT useful for diagnosing and producing vaccines against bacterial
PT infections such as otitis media and pneumonia

XX Claim 1; Page 67; 86pp; English.

XX The present sequence is that of BASB13 protein from Moraxella
CC catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis
CC media in children and pneumonia in adults. The invention provides
CC BASB13 polypeptides, and polynucleotides encoding them, as well as
CC expression vectors, host cells and methods for producing BASB13
CC polypeptides using recombinant methods. Also claimed is a vaccine
CC composition comprising a BASB13 polypeptide, an immunogenic

CC Fragment of a BASB13 polypeptide, or a polypeptide having at least
CC 85% amino acid sequence identity to BASB13, or comprising a
CC polynucleotide encoding such a polypeptide. A claimed method of
CC diagnosing a Moraxella infection involves identifying a BASB13
CC polypeptide or antibody. A claimed therapeutic composition useful
CC in treating humans with M. catarrhalis infection comprises at least
CC 1 antibody directed against a BASB13 polypeptide. BASB13
CC polypeptides also have utility in raising specific antibodies.
XX and in screening for antibacterial drugs.

Query Match 12.9%; Score 108; DB 22; Length 224;
Best Local Similarity 32.4%; Pred. No. 0.0028;
Matches 36; Conservative 16; Mismatches 35; Indels 24; Gaps 5;

OY 8 GSSLIIIVETVGCANFRQEVGATG-----AVGVGAGQLFGKSGRYAMAIG 59
Db 7 GVVLLASSMALGANTGT---TGNGTREGANVNRKAVIGAVAGAL---GCTATSKATGC 60

OY 60 -----AVLGGLISKTGSMDDQDKIKLNQSLKRYKAGQYTRMRNPDTGN 104
Db 61 EKTGRDAITLGAAGAAGAYMERQAK-----QTEQMGQGTGVTYHDTGN 107

RESULT 7

AAV34487 standard; Protein; 223 AA.

XX AAV34487;

XX 25-AUG-1999 (first entry)

XX Porphyromonas gingivalis protein PG3.

XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine; antigenic.

XX Porphyromonas gingivalis.

XX WO9929870-A1.

XX 17-JUN-1999.

XX 10-DEC-1998; 98MO-AU01023.

XX 04-AUG-1998; 98AU-0005028.

XX 31-DEC-1997; 97AU-0000839.

XX 30-JAN-1997; 97AU-0001182.

XX 10-MAR-1998; 98AU-0001546.

XX 09-APR-1998; 98AU-0002264.

XX 23-APR-1998; 98AU-0002911.

XX 05-MAY-1998; 98AU-0003128.

XX 22-MAY-1998; 98AU-0003358.

XX 29-JUL-1998; 98AU-0003654.

XX (CSLC-) CSL LTD.

XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA.

XX Ross BC, Rothel LJ, Webb EA.

XX WPI: 1999-385613/32.

XX N-PSDB; AAX91705.

XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis

XX Claim 1; Page 469; 588pp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the

30 VGAAATGAVVGGVAGQLEFGKSGRVAMATGCAVLGGTIGSKIGQSMDDQDKIKL 82

Db 62

VGTLCGAGALGAGTIGGAGGAGTAATVGAIGAGIAGSKIEFKMSQVNGAEL 114

RESULT 10

ABG15906

ID ABG15906 standard; Protein; 309 AA.

AC ABG15906;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #15897.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YN;

DR WPI: 2001-639362/73.

XX N-PSDB; AAS80093.

PT New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity

PS Claim 20; SEQ ID No 46265; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags (ESTs). (II) is useful for generating antibodies against it, detecting or imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations and to produce other types of data and products dependent on DNA and amino acid sequences. Ab000010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 309 AA:

Query Match

Best Local Similarity 11.7%; Score 98; DB 22; Length 309;

Matches 33; Conservative 11; Mismatches 39; Indels 14; Gaps 3;

DB 13 IISVFLVGCANFRS-----DEVAATGAAVGVAGOLFGKG--SGRVAMATGGAIVLGL 65
185 IISIPAVAVDNGKRNKMAQAFGALIGAVAGGVIGHNNGSSNGTAYGAVGVGAACAA 244

OY 66

IGSKI-----GOSMDQDKIKINOSLEKRAQOYT 95

DB 245 AGSMVNDKTLMEGVSLLTYKESTKYTSTGVKGFQFT 281

RESULT 11

AAB82611

ID AAB82611 standard; Protein; 528 AA.

AC AAB82611;

DT 02-OCT-2001 (first entry)

DE Spider recombinant silk protein PETNCS.

KW Spider; orb-weaver; silk protein; PETNCS; structural protein;

XX purification; fibre; spinning.

OS Nephila clavipes.

PN Key Location/Qualifiers

FT Misc-difference 417 /note= "encoded by TAT"

FT Misc-difference 427 /note= "encoded by CGA"

FT Misc-difference 522 /note= "encoded by GAG"

XX WO20015333-A1.

PD 26-JUL-2001.

PF 01-NOV-2000; 2000WO-US30086.

XX 20-JAN-2000; 2000US-0490291.

PA (MELT/) MELLO C M.

PA (ARCT/) ARCIDIACONO S.

PA (BUTL/) BUTLER M M.

PA (USSA) US SEC OF ARMY.

PI Mello CM, Arcidiacono S, Butler MM;

DR WPI: 2001-483136/52.

XX N-PSDB; AAB26304.

Recovering structural polypeptides in a biological sample, useful for purifying and spinning spider silks and other structural proteins, comprises treating the sample containing the polypeptides with an acid

Claim 2; Page 41-42; 49pp; English.

The present sequence is that of the orb-weaver spider (Nephila clavipes) recombinant silk protein PETNCS. The invention provides methods for purifying and spinning spider silks and other cells or other biological samples (such as non-recombinant derived cells), and enrich the purity and yields of structural proteins by hydrolysing many of the macromolecules while leaving the structural proteins intact. In the case of silk proteins, the resulting lysate is further purified by ion-exchange or affinity chromatography and processed into an aqueous-based mixture for into vector PET24 for recombinant expression in *Escherichia coli*, containing denaturant (3 M guanidine-HCl) and by affinity chromatography on nickel-NTA agarose. Products obtained using the methods of the invention can be used in the construction of many materials including films, fibres, woven articles, sutures, the ballistic protection, parachutes and parachute cords. The new method has the following advantages over prior art: it involves fewer steps, requires less time and smaller volumes of reagents.

CC results in better recovery of protein at higher purity (70-99%),
CC is easy to scale up, and the fibres are spun in an environmentally
CC benign solution reducing hazardous waste accumulation and cost.

SO Sequence 528 AA;

Query Match 10.9%; Score 91; DB 22; Length 528;
Best Local Similarity 51.1%; Pred. No. 0.51;
Matches 24; Conservative 2; Mismatches 19; Indels 2; Gaps 1;

OY 28 QEVGATGAVGVAGQ--LFGKSGRVAMAGAVLGILGSKIGQ 72
DB 320 QGAGAAAAAAGAGAGCGGCGAGAAAAAAGAGAGCGGCGAGQ 366

RESULT 12

ID ABB66232 standard; Protein; 2309 AA.

XX ABB66232;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 25488.

KM Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

DR N-PSDB: ABL10335.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Disclosure; SEQ ID NO 25488; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB101840-AB16175).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2309 AA;

Query Match 10.3%; Score 86.5; DB 22; Length 2309;
Best Local Similarity 27.4%; Pred. No. 9.9;
Matches 34; Conservative 18; Mismatches 55; Indels 17; Gaps 5;

OY 23 QNFSRQEVGAATGA---VVGVAAGLFGKSGRVAMAGAVLGILGSKIGSMDDQDK 79
DB 2124 QQQQQRQVGGNGPSPMNLGGRSGAVGSGSGN-----GGGGGGGAGGQVGGNGGNGV 2178

OY 80 IKLNQSLKXKAGQVTRWRNP-DTGNSTSYEPVTRQRYNKKQERROOYCFEQKAMIAQ 138
DB 2179 GSVGQS-----GGGQGRTPPIORPNMYPQHP-----QQQQQQQQQQQQREQAAAAVAA 2230

OY 139 QKOE 142
DB 2231 QRAQ 2234

RESULT 13

ID ABB58019 standard; Protein; 666 AA.

XX ABB58019;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 849.

KM Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

DR N-PSDB: ABL02122.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Disclosure; SEQ ID NO 849; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB101840-AB16175).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 666 AA;

Query Match 9.8%; Score 82; DB 22; Length 666;
Best Local Similarity 22.1%; Pred. No. 6.1;
Matches 30; Conservative 22; Mismatches 58; Indels 26; Gaps 2;

OY 31 GAATGAVGVAGQDLFGKSGRVAMAGAVLGILGSKIGSMDDQDKIKLNQSLKXK 90
DB 463 GPGTGGAGGGGSSSSNGAGTAAVS-GSVTTAGAGSGTGSNSNSNSGSGSVGSV 521
OY 91 AGQVTRWRNPDTGNSTSYV-----EPVTRQRYNKKQERROO 125
DB 522 SGSGSHPGTPTSLHAHSANGTSSSLGGLHLATPHQMVAAAGSPVVLHQQQQQQQHQQQ 581

OY 126 YCREFOQKAWIAGOK 141
 Db 582 Q0000000H0Q000000 597

RESULT 14

AAAY40097

ID AAY40097 standard; protein; 651 AA.

AC AAY40097;

DT 19-NOV-1999 (first entry)

DE Spider silk protein spidroine major 1.

KW Spider silk protein; spidroine major 1; cosmetic; make-up;

KW dermatological compositions; hair care; skin care; sunscreen;

OS hormone; moisturizer; skin disorder; skin care; sunscreen;

OS Nephila clavipes.

XX

PN FR2774588-A1.

PD 13-AUG-1999.

XX

PF 11-FEB-1998; 98FR-0001614.

PR 11-FEB-1998; 98FR-0001614.

XX (OREA) L'OREAL SA.

PA

PI Philippe M, Garson JC, Arraudeau JP;

XX WPI; 1999-510729/43.

DR

XX

PT Cosmetic or dermatological composition containing spider silk protein,

XX for hair or skin care, in make-up or sunscreens.

XX Claim 3; Fig 1; 32pp; French.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

N.clavipes dragline silk protein-1.

protein superfbre; major ampullate silk; orb web spider.

Nephila clavipes.

EP452925-A.

23-OCT-1991.

18-APR-1991; 91EP-0106217.

20-APR-1990; 90US-0511792.

(UYW-) UNIT OF WYOMING.

Lewis RV, Xu M, Hlman M;

WPI; 1991-312199/43.

N-PSDB; AAQ14183.

DNA encoding spider silk protein-1 and 2 and variants - isolated

from Nephila clavipes, for prodn. of spider silk protein and

fibres having desired characteristics

Claim 15; Page 23; 48pp; English.

The spider silk protein contains a basic 34 amino acid repeat. The

repeat itself contains 3 regions. The first comprises 0-9 amino

acids with a sequence AGR(GGX)2. This region is not highly

conserved. The second region has a sequence GAG(Axx) which is highly

conserved and is 8-10 amino acids long. The third segment is (GGX)5

and is 15 amino acids long and is very highly conserved. In most

cases x is A,Q, Y or L. Removal of the poly-(Ala) segments results

in a silk having lower elasticity.

Sequence 718 AA;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 15:09:08 : Search time 962.107 Seconds
(without alignments)
6775.784 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483

Sequence: 1 atcgctgctgcctgcacgag.....agggtattagacacgaaaaa 483

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

EST:*
1: em_estba:*
2: em_estbun:*
3: em_estln:*
4: em_estlmu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_huv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	8.7	238	10	BG240789 OV1_38_C0
2	42	8.7	402	10	BG356895 OV2_11_D0
3	42	8.7	419	10	BG356895 OV2_11_D0
4	42	8.7	428	10	BG356894 DGI_11_E1
5	42	8.7	446	10	BG176742 EM1_4_F01
6	42	8.7	500	10	BG052290 RH12_12
7	42	8.7	507	10	BG586874 FM1_31_F0
8	42	8.7	531	10	BG053591 RH12_11
9	42	8.7	531	10	BG053014 RH12_16
10	42	8.7	537	10	BG13844 EM1_20_C0
11	42	8.7	577	10	BG587805 FM1_40_G0
12	42	8.7	582	10	BG673858 EM1_20_C0
13	42	8.7	591	10	BG356763 DGI_12_B1
14	42	8.7	597	10	BG356800 DGI_12_E1
15	42	8.7	598	9	BG102589 RH12_34
16	42	8.7	600	10	AW672446 LG1_359_D
17	42	8.7	620	10	BG360868 DGI_67_A0
					BF481524 FM1_21_B0

18	42	8.7	649	10	BE355895 DGI_11_E1
19	40.4	8.4	548	10	BM328366 PIC1_28_D
20	40	8.3	513	9	AV430609 AV430609
21	40	8.3	768	12	AQ690676 nbx00082B
22	38	7.9	910	12	CNS0060N
23	38	7.9	1101	12	CNS0100Y
24	37.8	7.8	595	10	BI607335
25	37.8	7.8	826	10	BI522679
26	37.8	7.8	925	12	CNS0091P
27	37.6	7.8	1344	10	BE364680
28	37.4	7.7	350	10	BG052289 RH12_12
29	37.4	7.7	1101	12	CNS017SY
30	37.2	7.7	514	10	BG320949
31	37	7.7	908	10	BF253465
32	36.8	7.6	446	9	AU182544
33	36.8	7.6	1101	9	AL513871
34	36.4	7.5	1043	10	BG104511
35	36.2	7.5	595	12	BM478914
36	36.2	7.5	647	10	BJ003370
37	36.2	7.5	663	10	BJ003303
38	36.2	7.5	789	9	AU125061
39	36	7.5	1058	10	BG295311
40	35.8	7.4	217	10	BM445199
41	35.8	7.4	682	10	BE593319
42	35.6	7.4	467	9	AU182484
43	35.6	7.4	586	10	BF978932
44	35.6	7.4	653	10	BE293240
45	35.6	7.4	685	9	AU130542

ALIGNMENTS

RESULT 1
LOCUS BG240789 238 bp mRNA linear EST 15-FEB-2001
DEFINITION OV1_38.C06.g1.A002 ovary 1 (OV1) Sorghum bicolor cDNA, mRNA

ACCESSION BG240789
VERSION BG240789.1 GI:12775862
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt L.H.

TITLE An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805

Seq primer: PolyaTmX
High quality sequence start: 3
High quality sequence stop: 230
POLYA-NO.

FEATURES
source location/Qualifiers
1..238

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made

BASE COUNT 63 a 54 c 70 g 51 t
 FROM poly-A RNA in the cloning vector lambda ZAP II.
 Clones to be sequenced were prepared by mass excision."

Query Match
 Best Local Similarity 8.7%; Score 42; DB 10; Length 238;
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGAGAACTTACGCCGAGAGTTGGCGGCGCACCGGTCGCTGTGTGGCGGCTGT 119
 DB 79 CGCGGAGAACTTACGCCGAGAGTTGGCGGCGCACCGGTCGCTGTGTGGCGGCTGT 119
 QY 120 TGGCGGCGGCTGTGTGGCGGCGCACCGGTCGCTGTGTGGCGGCTGTGTGGCGGCTGT 138
 DB 139 CGACGCGGCGGCTGTGTGGCGGCGCACCGGTCGCTGTGTGGCGGCTGTGTGGCGGCTGT 179
 QY 180 TC 181
 DB 199 CC 200

RESULT 2
 BG356895 402 bp mRNA linear EST 06-MAR-2001
 LOCUS OY2.11.D01.g1.A002 Ovary 2 (OY2) Sorghum bicolor cDNA, mRNA
 DEFINITION
 ACCESSION BG356895
 VERSION BG356895.1 GI:13238881
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Sorghum.
 REFERENCE 1 (bases 1 to 402)
 AUTHORS Cordomier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 TITLE An EST database from Sorghum: ovaries of varying immature stages
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordomier-Pratt MM
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmp@prattuga.edu
 Seq primer: PolyTMix
 High quality sequence start: 47
 High quality sequence stop: 391
 POLYA-NO.

FEATURES

Location/Qualifiers

1..402
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /note="Organ: Ovary 2 (OV2)"
 from 8-week-old plants; Vector: pBluescript II from lambda
 Zap II; Site:1: XhoI; Site:2: EcoRI; The library was made
 from poly-A RNA in the cloning vector lambda Zap II.
 Clones to be sequenced were prepared by mass excision."
 BASE COUNT 107 a 82 c 117 g 96 t
 ORIGIN

Query Match
 Best Local Similarity 8.7%; Score 42; DB 10; Length 402;
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGAGAACTTACGCCGAGAGTTGGCGGCGCACCGGTCGCTGTGTGGCGGCTGT 119
 DB 140 CGCGGAGAACTTACGCCGAGAGTTGGCGGCGCACCGGTCGCTGTGTGGCGGCTGT 119
 QY 120 TGGCGGCGGCTGTGTGGCGGCGCACCGGTCGCTGTGTGGCGGCTGTGTGGCGGCTGT 138
 DB 139 CGACGCGGCGGCTGTGTGGCGGCGCACCGGTCGCTGTGTGGCGGCTGTGTGGCGGCTGT 179
 QY 180 TC 181
 DB 199 CC 200

DB 200 CGACGCGGCGGCTGTGTGGCGGCGCACCGGTCGCTGTGTGGCGGCTGTGTGGCGGCTGT 259
 QY 180 TC 181
 DB 260 CC 261

RESULT 3
 BE355894 419 bp mRNA linear EST 20-JUL-2000
 LOCUS DGL.11.E11.g1.A002 Dark grown 1 (DGL) Sorghum bicolor cDNA, mRNA
 DEFINITION
 ACCESSION BE355894
 VERSION BE355894.1 GI:9297451
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Sorghum.
 REFERENCE 1 (bases 1 to 419)
 AUTHORS Cordomier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 TITLE An EST database from Sorghum: dark-grown seedlings
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordomier-Pratt MM
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmp@prattuga.edu
 Seq primer: PolyTMix
 High quality sequence start: 16
 High quality sequence stop: 336
 POLYA-NO.

FEATURES

Location/Qualifiers

1..419
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /note="Organ: 5-day-old dark-grown seedlings; Vector: lambda Zap; Site:1: XhoI; Site:2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II.
 Clones to be sequenced were prepared by mass excision."
 BASE COUNT 102 a 82 c 116 g 119 t
 ORIGIN

Query Match
 Best Local Similarity 8.7%; Score 42; DB 10; Length 419;
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGAGAACTTACGCCGAGAGTTGGCGGCGCACCGGTCGCTGTGTGGCGGCTGT 119
 DB 78 CGCGGAGAACTTACGCCGAGAGTTGGCGGCGCACCGGTCGCTGTGTGGCGGCTGT 119
 QY 120 TGGCGGCGGCTGTGTGGCGGCGCACCGGTCGCTGTGTGGCGGCTGTGTGGCGGCTGT 137
 DB 138 CGACGCGGCGGCTGTGTGGCGGCGCACCGGTCGCTGTGTGGCGGCTGTGTGGCGGCTGT 179
 QY 180 TC 181
 DB 198 CC 199

RESULT 4
 BE176742 428 bp mRNA linear EST 31-OCT-2000
 LOCUS EM1.4.F01.g1.A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
 DEFINITION

sequence.
 accession BFL76742
 version BFL76742.1 GI:11064652
 keywords EST.
 source Sorghum.
 organism Sorghum bicolor
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 428)
 Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
 An EST database from Sorghum: developing embryos
 Unpublished (2000)
 contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmp@atluga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Pined quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: PolyTmix
 High quality sequence start: 32
 High quality sequence stop: 427
 POLYA-No.
 Location/Qualifiers
 1..428
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Embryo 1 (EM1)"
 /note="Organ: Embryos germinated for 24 hr. Vector:
 Bluescript II from Lambda zap II; Site_1: XhoI; Site_2:
 EcoRI; The library was made from poly-A RNA in the cloning
 vector lambda zap II. Clones to be sequenced were
 prepared by mass excision."
 BASE COUNT 103 a 80 c 119 g 126 t
 ORIGIN
 Query Match 8.7%; Score 42; DB 10; Length 428;
 Best Local Similarity 59.0%; Pred. No. 0.25;
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 QY 60 CGCCGAGACTTACGCCGAGAGATTGGCGCGCCACCGGTCGGTGGCGGTG 119
 || ||||| || || ||||| || || || ||||| || || || |||||
 Db 58 CGCGGAGAGACTGACCCGAGAGAGTTCGACGAGATGATCCGTAGCGTCGATGG 117
 QY 120 TGGCGGCGACCTGTTCGTAAGGCTCTGGCTGTGTCGATGCGGCATCGCGGTGGCGGT 179
 || ||||| || || ||||| || || ||||| || || ||||| || || |||||
 Db 118 CGAGCGCCAGATCAACTATGAGAGTTTGTAAAGTTATGATGCGCAAGTGAAGCGCGT 177
 QY 180 TC 181
 Db 178 CC 179
 RESULT 5
 BG052290 446 bp mRNA linear EST 25-JAN-2001
 LOCUS RH122_12_G04.g1_A003 Rhizome2 (RH122) Sorghum prolinquum cDNA, mRNA
 DEFINITION
 sequence.
 accession BG052290
 version BG052290.1 GI:12506830
 keywords EST.
 source Sorghum prolinquum.
 organism Sorghum prolinquum.
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 446)
 Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
 ,L.H.

TITLE
 JOURNAL
 COMMENT
 An EST database from Sorghum: Sorghum prolinquum rhizomes
 Unpublished (2000)
 contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmp@atluga.edu
 Seq primer: PolyTmix
 High quality sequence start: 11
 High quality sequence stop: 440
 POLYA-No.
 Location/Qualifiers
 1..446
 /organism="Sorghum prolinquum"
 /db_xref="taxon:132711"
 /clone_lib="Rhizome2 (RH122)"
 /note="Organ: Rhizomes; Vector: Bluescript II from Lambda
 zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
 from poly-A RNA in the cloning vector lambda zap II.
 Clones to be sequenced were prepared by mass excision."
 BASE COUNT 101 a 81 c 124 g 140 t
 ORIGIN
 Query Match 8.7%; Score 42; DB 10; Length 446;
 Best Local Similarity 59.0%; Pred. No. 0.25;
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 QY 60 CGCCGAGACTTACGCCGAGAGATTGGCGCGCCACCGGTCGGTGGCGGTG 119
 || ||||| || || ||||| || || || ||||| || || || |||||
 Db 51 CGCGGAGAGACTGACCCGAGAGAGTTCGACGAGATGATCCGTAGCGTCGATGG 110
 QY 120 TGGCGGCGACCTGTTCGTAAGGCTCTGGCTGTGTCGATGCGGCATCGCGGTGGCGGT 179
 || ||||| || || ||||| || || ||||| || || ||||| || || |||||
 Db 111 CGAGCGCCAGATCAACTATGAGAGTTTGTAAAGTTATGATGCGCAAGTGAAGCGCGT 170
 QY 180 TC 181
 Db 171 CC 172
 RESULT 6
 BF586874 500 bp mRNA linear EST 12-DEC-2000
 LOCUS Fm1_31_E06.g1_A003 Floral-Induced Meristem 1 (Fm1) Sorghum
 DEFINITION
 sequence.
 accession BF586874
 version BF586874.1 GI:11679198
 keywords EST.
 source Sorghum prolinquum.
 organism Sorghum prolinquum.
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 500)
 Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
 ,L.H.
 An EST database from Sorghum: floral-induced meristems
 Unpublished (2000)
 contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmp@atluga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Pined quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: PolyTmix
 High quality sequence start: 8
 High quality sequence stop: 495

FEATURES POLYA-No.
Location/Qualifiers
1. 500

BASE COUNT Clones to be sequenced were prepared by mass excision."
ORIGIN
1. 500
/organism="Sorghum prolinguam"
/db_xref="taxon:132711"
/clone_id="Rhizome2 (RH122)"
/note="Organ: Floral-induced Meristem 1 (FM1)"
pbluescript II from Lambda Zap II; Site: 1: XhoI; Site: 2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

Query Match
Best Local Similarity 59.0%; Score 42; DB 10; Length 500;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGAGAACTTCAGCCCGCCAGGAAGTTGGCGGCGGCGGTTGTGGCGGT 119
DB 89 CGCGGAGAACTTCAGCCCGCCAGGAAGTTGGCGGCGGCGGTTGTGGCGGT 119
QY 120 TGGCGGCGGCGGTTGTGGCGGCGGCGGCGGCGGTTGTGGCGGT 148
DB 149 CGACGGCCAGATCAACTATGAGAGTTTGTAAAGTTATGATGCGCAAGTGGAGCGGT 179
QY 180 TC 181
DB 209 CC 210

RESULT 7
BG053591
LOCUS 507 bp mRNA linear EST 25-JAN-2001
DEFINITION RH122_11.C03.g1.A003 Rhizome2 (RH122) Sorghum prolinguam cDNA, mRNA
ACCESSION BG053591
VERSION BG053591.1 GI:12509440
KEYWORDS EST.
SOURCE Sorghum prolinguam.
ORGANISM Sorghum prolinguam.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
1 (bases 1 to 507)
Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
L.H.
An EST database from Sorghum: Sorghum prolinguam rhizomes
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@pratt.uga.edu
Seq primer: T7
High quality sequence start: 95
High quality sequence stop: 507
POLYA=Yes.

FEATURES
source Location/Qualifiers
1. 507

/organism="Sorghum prolinguam"
/db_xref="taxon:132711"
/clone_id="Rhizome2 (RH122)"
/note="Organ: Rhizomes; Vector: pbluescript II from Lambda Zap II; Site: 1: XhoI; Site: 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II."

BASE COUNT Clones to be sequenced were prepared by mass excision."
ORIGIN
109 a 92 c 142 g 164 t

Query Match
Best Local Similarity 59.0%; Score 42; DB 10; Length 507;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGAGAACTTCAGCCCGCCAGGAAGTTGGCGGCGGCGGTTGTGGCGGT 119
DB 54 CGCGGAGAACTTCAGCCCGCCAGGAAGTTGGCGGCGGCGGTTGTGGCGGT 119
QY 120 TGGCGGCGGCGGTTGTGGCGGCGGCGGCGGTTGTGGCGGT 113
DB 114 CGACGGCCAGATCAACTATGAGAGTTTGTAAAGTTATGATGCGCAAGTGGAGCGGT 173
QY 180 TC 181
DB 174 CC 175

RESULT 8
BG053014
LOCUS 531 bp mRNA linear EST 25-JAN-2001
DEFINITION RH122_16.D04.g1.A003 Rhizome2 (RH122) Sorghum prolinguam cDNA, mRNA
ACCESSION BG053014
VERSION BG053014.1 GI:12508270
KEYWORDS EST.
SOURCE Sorghum prolinguam.
ORGANISM Sorghum prolinguam.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
1 (bases 1 to 531)
Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
L.H.
An EST database from Sorghum: Sorghum prolinguam rhizomes
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@pratt.uga.edu
Seq primer: PolyT/Mix
High quality sequence start: 42
High quality sequence stop: 527
POLYA=No.

FEATURES
source Location/Qualifiers
1. 531

/organism="Sorghum prolinguam"
/db_xref="taxon:132711"
/clone_id="Rhizome2 (RH122)"
/note="Organ: Rhizomes; Vector: pbluescript II from Lambda Zap II; Site: 1: XhoI; Site: 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 122 a 101 c 152 g 154 t
ORIGIN 2 others

Query Match
Best Local Similarity 59.0%; Score 42; DB 10; Length 531;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGAGAACTTCAGCCCGCCAGGAAGTTGGCGGCGGCGGTTGTGGCGGT 119
DB 121 CGCGGAGAACTTCAGCCCGCCAGGAAGTTGGCGGCGGCGGTTGTGGCGGT 119
QY 120 TGGCGGCGGCGGTTGTGGCGGCGGCGGCGGTTGTGGCGGT 180
DB 181 CGACGGCCAGATCAACTATGAGAGTTTGTAAAGTTATGATGCGCAAGTGGAGCGGT 240

OY 180 TC 181
DB 241 CC 242

RESULT 9
BG713844
LOCUS
DEFINITION BG713844 537 bp mRNA linear EST 08-MAY-2001
sequence.
EM1_20_C08_g2_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA

ACCESSION BG713844
VERSION BG713844
KEYWORDS GI:14007794
SOURCE EST.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 537)
Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
An EST database from Sorghum: developing embryos
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@prattuga.edu

COMMENTS Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: PolyTmix
High quality sequence start: 62
High quality sequence stop: 536
POLYA-No.

FEATURES
Location/Qualifiers
source 1..537
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr: Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 118 a 98 c 152 g 169 t
ORIGIN

Query Match 8.7%; Score 42; DB 10; Length 537;
Best Local Similarity 59.0%; Pred. No. 0.27;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 60 CGCCGAACTTCAGCGCCGAGAGTTGGCGGCGGCGGCGGCGGT 119
DB 92 CGCGGAAAGCTACCGACGAGAGTCCGAGATGATCGTGAAGCTACGTCATGG 151
OY 120 TGGCGGCGACGCTTGGTAAAGGCTCTGCTCGTGCATGCGCATCCGCGGTGG 179
DB 152 CGACGGCGACGATCACTATGAAGATTGTTAAGCTTATGATGCGCAAGTGAAGAGCGGT 211
OY 180 TC 181
DB 212 CC 213

RESULT 10
BF587805
LOCUS
DEFINITION BF587805 561 bp mRNA linear EST 12-DEC-2000
propinquum cDNA, mRNA sequence.

VERSION BF587805.1 GI:11680115
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 561)
Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt,L.H.
An EST database from Sorghum: floral-induced meristems
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@prattuga.edu

COMMENTS Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: PolyTmix
High quality sequence start: 81
High quality sequence stop: 561
POLYA-No.

FEATURES
Location/Qualifiers
source 1..561
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Floral-Induced Meristem 1 (FMI)"
/note="Organ: Floral-Induced Meristems; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 140 a 111 c 160 g 150 t
ORIGIN

Query Match 8.7%; Score 42; DB 10; Length 561;
Best Local Similarity 59.0%; Pred. No. 0.28;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 60 CGCCGAACTTCAGCGCCGAGAGTTGGCGGCGGCGGCGGCGGT 119
DB 202 CGCGGAAAGCTACCGACGAGAGTCCGAGATGATCGTGAAGCTACGTCATGG 261
OY 120 TGGCGGCGACGCTTGGTAAAGGCTCTGCTCGTGCATGCGCATCCGCGGTGG 179
DB 262 CGACGGCGACATCACTATGAAGATTGTTAAGCTTATGATGCGCAAGTGAAGAGCGGT 321
OY 180 TC 181
DB 322 CC 323

RESULT 11
BG673858
LOCUS
DEFINITION BG673858 577 bp mRNA linear EST 01-MAY-2001
sequence.
EM1_20_C08_g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA

ACCESSION BG673858
VERSION BG673858
KEYWORDS GI:13919511
SOURCE EST.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

REFERENCE 1 (bases 1 to 577)
 AUTHORS Reid, S.P., Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.
 TITLE An EST database from Sorghum: developing embryos
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu

Clade: Panicoideae; Andropogoneae; Sorghum.
 Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
 Seq primer: PolyTMix
 High quality sequence start: 57
 High quality sequence stop: 572
 POLYA-No.

FEATURES

source location/Qualifiers
 1..577
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Embryo 1 (EM1)"
 /note="Organ: Embryos germinated for 24 hr; Vector: Bluescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
 BASE COUNT 129 a 106 c 166 g 176 t
 ORIGIN

Query Match
 Best Local Similarity 8.7%; Score 42; DB 10; Length 577;
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

60 CGCCGCAACTTCAGCCGCGAGAGTTGGCGCGCCGCGGTGTGTGGCGGTGT 119

121 CGCGGAGAGTGCACGACGAGAGGTCGACGAGATGTCGTGAGCGCTGCGATGG 180

120 TCCCGCGCAGCTGTGGTAAAGCTCTGCTGTGTCGATGCGCATCGCGTGGGT 179

181 CGACGCGCCAGATCACTATGAGAGTTTGTAAAGTTATGATGCCAAGTAGAGCGGT 240

180 TC 181

241 CC 242

RESULT 12
 BE356763 582 bp mRNA linear EST 20-JUL-2000

LOCUS DGI_12.B12.g1_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA

ACCESSION BE356763
 BE356763
 BE356763.1 GI:9298320

VERSION BE356763.1
 BE356763.1 GI:9298320

KEYWORDS EST.
 EST.

SOURCE sorghum.
 sorghum.

ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

REFERENCE 1 (bases 1 to 582)
 Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt, L.H.

AUTORS An EST database from Sorghum: dark-grown seedlings
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM

DEPARTMENT OF Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805

FEATURES
 source location/Qualifiers
 1..591
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"

Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
 Seq primer: PolyTMix
 High quality sequence start: 6
 High quality sequence stop: 553
 POLYA-No.

FEATURES

source location/Qualifiers
 1..582
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Dark Grown 1 (DGI)"
 /note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
 BASE COUNT 145 a 116 c 166 g 155 t
 ORIGIN

Query Match
 Best Local Similarity 8.7%; Score 42; DB 10; Length 582;
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

60 CGCCGCAACTTCAGCCGCGAGAGTTGGCGCGCCGCGGTGTGTGGCGGTGT 119

213 CGCGGAGAGTGCACGACGAGAGGTCGACGAGATGTCGTGAGCGCTGCGATGG 272

120 TCCCGCGCAGCTGTGGTAAAGCTCTGCTGTGTCGATGCGCATCGCGTGGGT 179

273 CGACGCGCCAGATCACTATGAGAGTTTGTAAAGTTATGATGCCAAGTAGAGCGGT 332

180 TC 181

333 CC 334

RESULT 13
 BE356800 591 bp mRNA linear EST 20-JUL-2000

LOCUS DGI_12.E10.g1_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA

ACCESSION BE356800
 BE356800
 BE356800.1 GI:9298357

VERSION BE356800.1
 BE356800.1 GI:9298357

KEYWORDS EST.
 EST.

SOURCE sorghum.
 sorghum.

ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

REFERENCE 1 (bases 1 to 591)
 Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt, L.H.

AUTORS An EST database from Sorghum: dark-grown seedlings
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM

DEPARTMENT OF Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: PolyTMix
 High quality sequence start: 10
 High quality sequence stop: 557
 POLYA-No.

FEATURES
 source location/Qualifiers
 1..591
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"

```

/clone_lib="Dark Grown 1 (DG1)"
/notes="Organ: 5-day-old dark-grown seedlings; Vector:
lambda zap; Site_1: XhoI; Site_2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda Zap II
Clones to be sequenced were prepared by mass excision."
BASE COUNT      146 a      114 c      168 g      163 t
ORIGIN

```

Query Match	8.7%	Score 42;	DB 10;	Length 591;
Best Local Similarity	59.0%;	Pred. No. 0.29;		
Matches	72;	Conservative	0;	Mismatches 50;
			Indels	0;
			Gaps	0;

Oy 60 CGCCAGACTTCAGCCGCCAGGAATTGGCGCCGCTGTGTTGGCGGTG 119
 | | | | | | | | | | | | | | | | | |
Db 210 CCGCGAGAACTGACCGACGAGAGTTCGACGAGATGATCCGTGAGCGCTGACGTGATGG 269

Oy 120 TGCGGGCCACCTGTTCCGTAAAGGCCTCTGGTCGTGTGCATGCCCATCGGCGGTGGCGT 179
| | | | | | | | | | | | | | | | | |
Db 270 CGACGGCCACATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAAGTGAGGAGCGGT 329

QY	180 TC	181
Db	330 CC	331

RESULT_14	LOCUS	DEFINITION
BG102589	597 bp	mRNA linear EST 30-JAN-2001
RIH122_34_H05_g1_A003	Rhizome2 (RIH122)	Sorghum propliquum cDNA, mRNA

VERSION	BG102589.1	GI:12617422
KEYWORDS	EST.	
SOURCE	Sorghum propinquum.	

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE
AUTHORS
1 (bases 1 to 597)
Cordonier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt, L.H.

TITLE	An EST database from Sorghum: Sorghum propinquum rhizomes
JOURNAL	Unpublished (2000)
COMMENT	Contact: Cordonnier-Pratt MM

The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: +1 706 542 2020

Email: mmpratt@cuga.edu

High quality sequence start: 5933
High quality sequence stop: 5933
POLYA-No.

FEATURES	Location/Qualifiers
source	1. .597

```

/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizomes (RHL22)"
/notes="Organ: Rhizomes: plasmidscript II from Lambda
zap II: Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT      149 a      130 c      170 g      148 t
ORIGIN

```

Query Match	8.78; Score 42; DB 10; Length 597;
-------------	------------------------------------

Matches	72;	Conservative	0;	Mismatches	50;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

Oy 60 CGCCGAGAACTTTCACGCCCGCCACGGAATTGGCGCGGCCACC GGTTGGTGCGGTGT 119
||| ||| | | | | | | | | | | | |
Db 267 CGCGGAGAAGCTGACCGACGAGAGGTCGACGAGATGATCCGTGAGGCTGCACGTCCAATGG 326

Oy	120	TGCGCCAGCTGTTCGGTAAAGCCTCGTGCTGCATGGCAACCCGCGGCGGT	179
Dd	327	CGACGCCAAGTCAACTATGAAGAGTTTGTTAAGCTATATCATGGCCAAGTAGAGAGCGGT	386
Oy	180	TTC 181	
Dd	387	CC 388	

RESULT 15
AM672446

M002944C F00 1-
END 1-1-1980 PGE 10 TRF 00000

LOCUS	588 bp	mrna	linear	EST 19-JUL-200
DEFINITION	LG1_359_D09.g1_A002	Light	Grown	1 (LGI) Sorghum bicolor CDNA, mRNA
SEQUENCE	sequence.			

VERSION	AW672446.1	GI:7536365
KEYWORDS	EST.	
SOURCE	sorghum.	

REFERENCE

Chen, S. C. and W. C. Chen. 1973. New records of the plant kingdom of Taiwan. *Acta Botanica Sinica* 1: 1-11.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum. 1 (bases 1 to 598)

TITLE	An EST database from Sorghum: light-grown seedlings
JOURNAL	Unpublished (2000)
COMMENT	Contact: Cordonnier-Pratt MM Department of Botany

Tel: 706 542 1860
Fax: 706 542 1805
Email: mpmratt@uga.edu

below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: T7

High quality sequence start: 71
High quality sequence stop: 598
POLYA=Yes.

FEATURES	Location/Qualifiers
source	1. 598
	/organism="Sorghum bicolor"

```

/clone_lib="Light Grown 1 (LG1)"
/note="Organ: 10- to 14-day-old light-grown (greenhouse)

```

vector lambda

BASE COUNT	151 a	128 c	173 g	146 t
ORIGIN				

Query Match	8.78;	Score 42;	DB 9;	Length 598;
Best Local Similarity	59.08;	Pred. NO.	0.29;	

	Matches	72:	Conservative	0:	Mismatches	50:	Indels	0:	Gaps
Q7	60	CGCCCAAGACTTCAGGCCCGCCGGAAGTGGGGCGGCCCGCCGTCGGTGTGGGGCGGT	11.9						
Db	279	CGCGCAAGAGCTGACCCAGCGAGGGGTGACGACAAATGCTTCGTAAGCTGACGTCAATGG	338						
Qy	120	TTCGGCGCAGCTGTTCCGTTAAGGCTGTGCTGTGTGCATGCGCATCGGCGTGCAGT	17.9						
Db	339	CGACGGCCAGATTCAACTGTGAAGAGTTGTTTAAGGTTATGATGTGCCAAGTCAGAGCGGT	398						

QY 180 TC 181

Db 399 CC 400

Search completed: October 27, 2002, 18:34:55
Job time : 966.107 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 16:06:08 ; Search time 35.3143 Seconds
(without alignments)
3359.570 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483
Sequence: 1 atgcgtgctgcctgcaggg.....agtgatcaccgcagaaaa 483

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/prodata/2/1na/5A.COMB.seq: *
2: /cgn2_6/prodata/2/1na/5B.COMB.seq: *
3: /cgn2_6/prodata/2/1na/6A.COMB.seq: *
4: /cgn2_6/prodata/2/1na/6B.COMB.seq: *
5: /cgn2_6/prodata/2/1na/6CTOS.COMB.seq: *
6: /cgn2_6/prodata/2/1na/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36.6	7.6	303	US-08-556-978B-80	Sequence 80, Appl
2	36.6	7.6	303	US-08-556-978B-81	Sequence 81, Appl
3	33.6	7.0	4403765	US-09-103-840A-2	Sequence 2, Appl
4	33.6	7.0	4411529	US-09-103-840A-1	Sequence 1, Appl
5	33.4	6.9	962	US-08-765-907A-16	Sequence 16, Appl
6	33.4	6.9	1052	US-08-403-852D-10	Sequence 10, Appl
7	33.4	6.9	1052	US-08-510-646B-10	Sequence 10, Appl
8	33.4	6.9	1052	US-09-231-818-10	Sequence 10, Appl
9	33.4	6.9	2888	US-08-765-907A-1	Sequence 1, Appl
10	33.2	6.9	23673	US-09-773-816-1	Sequence 1, Appl
11	33.2	6.9	4403765	US-09-103-840A-2	Sequence 2, Appl
12	33.2	6.9	4411529	US-09-103-840A-1	Sequence 1, Appl
13	33.2	6.8	1295	US-08-332-747-1	Sequence 1, Appl
14	32.4	6.7	535	5187077-4	Patent No. 5187077
15	32.4	6.7	535	5427925-3	Patent No. 5427925
16	32.4	6.7	599	5187077-24	Patent No. 5187077
17	32.4	6.7	599	5427925-22	Patent No. 5427925
18	32.4	6.7	714	5187077-16	Patent No. 5187077
19	32.4	6.7	1437	5427925-14	Patent No. 5427925
20	32.4	6.7	1437	5187077-16	Patent No. 5187077
21	32.4	6.7	1820	US-09-530-212A-1	Sequence 1, Appl
22	32.4	6.6	2365	US-08-363-208-1	Sequence 1, Appl
23	32.4	6.6	2365	US-09-137-478-1	Sequence 1, Appl
24	32.4	6.6	3300	US-08-194-279A-6	Sequence 6, Appl
25	32.4	6.6	3300	US-08-614-377A-6	Sequence 6, Appl
26	32.4	6.6	3300	US-09-142-648B-6	Sequence 6, Appl
27	31.8	6.6	13987	US-08-804-227C-13	Sequence 13, Appl

28	31.8	6.6	44377	2	US-08-804-227C-7	Sequence 7, Appl
29	31.8	6.6	44377	2	US-08-804-198-1	Sequence 1, Appl
30	31.6	6.5	5036	4	US-09-177-349-2	Sequence 2, Appl
31	31.2	6.5	7218	1	US-08-232-463-14	Sequence 14, Appl
32	30.8	6.4	1843	1	US-07-918-023-1	Sequence 1, Appl
33	30.8	6.4	2171	3	US-08-851-843A-100	Sequence 100, App
34	30.8	6.4	2171	4	US-08-974-549A-266	Sequence 266, App
35	30.8	6.4	2171	4	US-08-854-050-100	Sequence 100, App
36	30.8	6.4	2171	4	US-09-430-323-100	Sequence 100, App
37	30.6	6.3	882	4	US-08-818-112-138	Sequence 138, App
38	30.6	6.3	882	4	US-08-818-111-133	Sequence 138, App
39	30.6	6.3	882	4	US-09-056-556-138	Sequence 138, App
40	30.6	6.3	1689	1	US-08-442-884-2	Sequence 2, Appl
41	30.6	6.3	1820	1	US-08-173-508-7	Sequence 7, Appl
42	30.6	6.3	1821	2	US-08-265-310-7	Sequence 7, Appl
43	30.6	6.3	1821	3	US-08-951-742-7	Sequence 7, Appl
44	30.6	6.3	4030	4	US-09-293-505-1	Sequence 1, Appl
45	30.6	6.3	4705	1	US-07-998-003A-96	Sequence 96, Appl

ALIGNMENTS

RESULT 1
US-08-556-978B-80
Sequence 80, Application US/08556978B
Patent No. 6268169
GENERAL INFORMATION:
APPLICANT: FAHNESTOCK, STEPHEN F.
TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
TITLE OF INVENTION: SPIDER SILK ANALOGS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/556,978B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/077,600
FILING DATE: JUNE 15, 1993
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9389-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-556-978B-80

Query Match 7.6%; Score 36.6; DB 4; Length 303;
Best Local Similarity 57.4%; Pred. No. 0.043;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
85 GTTGGCGCCGACCGGTCGTTGTTGGCGGCGTTCGCGCAGCTGTTCCGTAAGGC 144
|||||

us-09-677-374-3.rni

Db	138	GCTCCGACAGCTGTTGGTCTGTGCAGGGCGCTTTGGCTACACAAAGGGCCGGTCAMGC	197
OY	145	TCGGTCGTGTGTCATGATGATCGATGGCGGTCGCTTCTGGAGGCGGTCTGATGAGCT	199
Db	198	GCTGGTCACGACGAGCTGCGCGTGGCGGTGCACAGGATGATATGTTGCT	252

```

RESULT 2
US-08-556-978B-81
; Sequence 81, Application US/0856978B
; Patent No. 6268169
GENERAL INFO:

```

NAME: FAHNESTOCK, STEPHEN F.
TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
TITLE OF INVENTION: SPIDER SILK ANALOGS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMMENTS:

COMPILED READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/556,978B
FILING DATE:
CLASS:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/077,600
FILING DATE: JUNE 15, 1993
ATTORNEY/AGENT INFORMATION:

NAME: FLOYD, LINDA AXAETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9389-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0111

INFORMATION FOR SEQ ID NO:4
 SEQUENCE CHARACTERISTICS: 81
 LENGTH: 303 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 8-556-978B-81

ery Match	7.6%;	Score 36.6;	DB 4;	Length 303,
st Local	Similarity	57.4%;		
tches	66;	Conservative	0.043;	

85 GTTGGCGGCGCCACCGCTGGGTGTGTGGCGGTCCTCCCC	Mismatches	0;	Indels	0;	Gaps
--	------------	----	--------	----	------

87 GCTGGCGGACAGCTGCTGTGTCGTGCATCAGCCGTTTGGCTAAAGCC 144
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
145 TCTGTGCTGTGTGTGATGCCCATCGCGGTGCTTGTGGGCGCTCAATTGCT 199
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

...GAGCTGGCCGTGCAGCCCAAGTGGATATGTGGCT 201

13
-103-840A-2/C
ence 2, Application US/09103840A
ent No. 6294328
RAL INFORMATION:
LICANT: FLEISCHMAN, Robert D.
LICANT: WHITE, Owen R.
LICANT: FRASER, Claire M.
LICANT: VENTER, John C.

```

: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: TITLE OF INVENTION: TUBERCULOSIS
: FILE REFERENCE: 24566-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

```

Query Match	7.08;	Score 33.6;	DB 4;	Length 4403765
Best Local Similarity	51.38;			
Matches	78;	Conservative		

	82	GAACTTGGCGCGCCACCGGTGCGGTTGTGGGCCcttccccccc	0	Gaps	0
QY		82	GAACTTGGCGCGCCACCGGTGCGGTTGTGGGCCcttccccccc	74;	Mismatches
				Indels	0;

Db 2884704 GATGTGTTGCCGCCGCCGGCGTGCCTCCGTCCTCAGCTTGCTAA 141

QY 142 GGCTGTGTCGTGTGCATGGCCATCGGCCGAGCCTTCCGTGCT 28

DbpA 2884644 GAGTGGGCCAGGCGCGCTTCTGGGCGCTGATTGGCTCT 201

202 AAATCGGTCAGAGCATGGACCAAGGCTGTTCCAGCTGTTGCCGTTGCTTCTGCAAA 2884585

db 2884584 AAAGCGTGGGCGCATCTTCAGATTA 23

...GCGTGGTTTGACCTTCGATGA 2884553

RESULT 4
S-09-103-840A-1/c

Sequence 1, Application US/09103840A
Patent No. 6294336

GENERAL INFORMATION:
APPLICANT: FERGUSON

APPLICANT: LEISCHMAN, Robert
WHITE, Owen R.
APPLICANT:

APPLICANT: VENTER, John C

TITLE OF INVENTION: DNA SEQUENCE
TITLE OF INVENTION: TUBERCULOSIS
TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: ITS 606 111

CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 3

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1

LENGTH: 441152
TYPE: DNA

ORGANISM: *Mycobacterium tuberculosis*
OTHER INFORMATION:

09-103-840A-1

```
Very Match      7.08: $$$$ ~
```

Local Similarity 48.9%; Pred. No. 18
atches 90; Conservative 0.00

	Mismatches	94;	Indels	0;	Gaps
39 CTCTGTTTTCAGTACCGCCGCCGCATC					

061470 CACTGCGGGCTTCCCTTTTGGCGGCCAC 98

99 CGGTTGGCGCTTCCGCCGTGCCTC
-----| | | | |
-----| | | | |

061410 GGGGCCCCCGGGTGTTCGCCAGCTGTTCCGTAAGGCTCTGCCTCCTGTC 158

[illegible][illegible]

CAGAGGAACGGCACCGGTGCCAAGGGCGGTGACGCGCGCTGCAGGTTCCCATTC

218

-----COUNCIL 2061291

Qy 219 GGAC 222
Db 2061290 GGTC 2061287

RESULT 5
US-08-765-907A-16/c
Sequence 16, Application US/08765907A
Patent No. 6352839
GENERAL INFORMATION:
APPLICANT: BLANC, Veronique
APPLICANT: THIBAUT, Denis
APPLICANT: BAMA-JACQUES, Nathalie
APPLICANT: BLANCHE, Francis
APPLICANT: COUZET, Joel
APPLICANT: BARRIERE, Jean-Claude
APPLICANT: DEBUSCHE, Laurent
APPLICANT: FAMECHON, Alain
APPLICANT: PARIS, Jean-Marc
APPLICANT: DUTRUC-ROSSET, Gilles
TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
FILE REFERENCE: Streptogramin genes
CURRENT APPLICATION NUMBER: US/08/765,907A
CURRENT FILING DATE: 1997-03-20
NUMBER OF SEQ. ID NOS: 17
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 16
LENGTH: 962
TYPE: DNA
ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-16

Query Match 6.9%; Score 33.4; DB 4; Length 962;
Best Local Similarity 55.7%; Pred. No. 0.68;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 75 CCGCAGAGATTGGCGCCGACCGGTGCGGTGCGGCGGTGTCGCCGCGAGCTGTT 134
Db 243 CGGCAGGAGAGGTGTGAGGTCGCCCGGGGAGCTGCGAGCGCGCGCAGC 184
Qy 135 CGGTAAAGCTCTGCTGTGTGATGCGCATGCGCGGTGCGGTGCGGCGGT 189
Db 183 GGGTGTGCGCCGCGGCGCGGTGCGCGGTGATCCGCGCGCGGTGACGTGCCCGGT 129

RESULT 6
US-08-403-852D-10/c
Sequence 10, Application US/08403852D
Patent No. 5891695
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanche, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Crey-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3115
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852D
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1052 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 84..962
OTHER INFORMATION: /product= "Gene Papw"
US-08-403-852D-10

Query Match 6.9%; Score 33.4; DB 2; Length 1052;
Best Local Similarity 55.7%; Pred. No. 0.71;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 75 CCGCAGAGATTGGCGCCGACCGGTGCGGTGCGGCGGTGTCGCCGCGAGCTGTT 134
Db 243 CGGCAGGAGAGGTGTGAGGTCGCCCGGGGAGCTGCGAGCGCGCGCAGC 184
Qy 135 CGGTAAAGCTCTGCTGTGTGATGCGCATGCGCGGTGCGGTGCGGCGGT 189
Db 183 GGGTGTGCGCCGCGGCGCGGTGCGCGGTGATCCGCGCGCGGTGACGTGCCCGGT 129

RESULT 7
US-08-510-646B-10/c
Sequence 10, Application US/08510646B
Patent No. 6077699
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanche, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Crey-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA

ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,646B
FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-01000
TELEPHONE: (202) 408-4400
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1052 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 84..962
OTHER INFORMATION: /product= "Gene Papw"

Query Match
Best Local Similarity 55.7%; Score 33.4; DB 3; Length 1052;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 75 CCCCAGAGAGTGGGCGGCCACCGGTCGAGTGTGGGCGGTGTCGCGCCAGCTGTT 134
DB 243 CCGCCAGCAGAGTGTGAGAGTCCCGGCGGACCTGGCAGGCGTGGCGCCAGCAGCC 184
QY 135 CCGTAAAGCTGCTGCTGTCGATGCGCATGCGCAGCGGTCGTCGAGCGGT 189
DB 183 GGGTGTGCGCCGCGGCGGTGATCCGCGCGCTGACCTGCTCCCGGT 129

RESULT 8
US-09-231-818-10/c

; Sequence 10, Application US/09231818
; Patent No. 6171846

; GENERAL INFORMATION:

; APPLICANT: Blanc, Veronique

; APPLICANT: Blanche, Francis

; APPLICANT: Crouzet, Joel

; APPLICANT: Jacques, Nathalie

; APPLICANT: Lacroix, Patricia

; APPLICANT: Thibaut, Denis

; APPLICANT: Zagorec, Monique

; APPLICANT: Debussche, Laurent

; APPLICANT: De Crecy-Lagard, Valerie

; TITLE OF INVENTION: Polypeptides Involved in The

; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences

; CODING FOR THESE POLYPEPTIDES AND THEIR USE

; APPLICANT: BARRIERE, Jean-Claude

; APPLICANT: DEBUSSCHE, Laurent

; APPLICANT: DEBUSSCHE, Laurent

; APPLICANT: DEBUSSCHE, Laurent

; APPLICANT: DEBUSSCHE, Laurent

NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESS: Finegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/231,818
FILING DATE:

PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852
FILING DATE: 10-MAY-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELEPHONE: (202) 408-4400
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1052 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 84..962
OTHER INFORMATION: /product= "Gene Papw"

Query Match
Best Local Similarity 55.7%; Score 33.4; DB 4; Length 1052;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 75 CCCCAGAGAGTGGGCGGCCACCGGTCGAGTGTGGGCGGTGTCGCGCCAGCTGTT 134
DB 243 CCGCCAGCAGAGTGTGAGAGTCCCGGCGGACCTGGCAGGCGTGGCGCCAGCAGCC 184
QY 135 CCGTAAAGCTGCTGCTGTCGATGCGCATGCGCAGCGGTCGTCGAGCGGT 189
DB 183 GGGTGTGCGCCGCGGCGGTGATCCGCGCGCTGACCTGCTCCCGGT 129

RESULT 9
US-08-765-907A-1/c

; Sequence 1, Application US/08765907A
; Patent No. 6352839

; GENERAL INFORMATION:

; APPLICANT: Blanc, Veronique

; APPLICANT: Thibaut, Denis

; APPLICANT: Bamas-Jacques, Nathalie

; APPLICANT: Blanche, Francis

; APPLICANT: Crouzet, Joel

; APPLICANT: BARRIERE, Jean-Claude

; APPLICANT: DEBUSSCHE, Laurent

; APPLICANT: DEBUSSCHE, Laurent

; APPLICANT: DEBUSSCHE, Laurent

; APPLICANT: DEBUSSCHE, Laurent

; APPLICANT: DEBUSSCHE, Laurent

; APPLICANT: DEBUSSCHE, Laurent

; APPLICANT: DEBUSSCHE, Laurent

; APPLICANT: DEBUSSCHE, Laurent

```
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Multisynthesis
; CURRENT APPLICATION NUMBER: US/08/765,907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-1
```

```
Query Match          6.9%; Score 33.4; DB 4; Length 2888;
Best Local Similarity 55.7%; Pred. No. 1.1;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
```

```
OY 75 CCGCCAGAGATTGGCCGCCACCGGTGCGGTTGGCGGTTGCCGCCAGCTGTT 134
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2418 CGGCCAGCGAGTGTGTCAGGTCCCGCGGCGACCTGGCAGCGCTGGCGCCAGCACC 2359
OY 135 CGGTAAGGCTGTGCTGTGTCGATGCGCATGCGGCGGTGCGGTGCGGCT 189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2358 GGGTGTGCGCCCGCGGCGGTGCGGTGATCCCGCGCGGTGCTGCGCCGCT 2304
```

```
RESULT 10
US-09-773-816-1
; Sequence 1, Application US/09773816
; Patent No. 6340774
; GENERAL INFORMATION:
; APPLICANT: Stanford University
; APPLICANT: Khosla, Chaitan
; TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR
; FILE REFERENCE: 28600-20210.00
; CURRENT APPLICATION NUMBER: US/09/773,816
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/243,458
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/179,305
; PRIOR FILING DATE: 2000-01-31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 23673
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(23623)
; OTHER INFORMATION: n = A,T,C or G
US-09-773-816-1
```

```
Query Match          6.9%; Score 33.2; DB 4; Length 23673;
Best Local Similarity 52.1%; Pred. No. 3.1;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
```

```
OY 55 GGTGGCCGCCAGAACTTCAGCCGCCAGAAATTGGCCGCCCGCGCGGTGTGGGC 114
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8366 GGGTTTCGGCAGATGACGTGCGCAGGAGGACGCCCGCGCGGTGTGGGT 8325
OY 115 GGTGTTCCGCGCAGTGTTCGTAAGGCTGTGCTGTGTCGATGCGCATCGGCGGT 174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8326 GGGTTCGCTGACACGACGATGTGTCACACAGCCGCGCTCTGTCAGTTGCCGAGGCGGC 8385
OY 175 GCGGTTCTGGCGGCTGTGATG 196
    ||||| ||||| ||||| |||||
DB 8386 GCTGCTCTTGCGCATCTGCATG 8407
```

```
RESULT 11
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```
Query Match          6.9%; Score 33.2; DB 4; Length 4403765;
Best Local Similarity 50.6%; Pred. No. 23;
Matches 80; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
```

```
OY 59 GCGCCAGAACTTCAGCCGCCAGAAATTGGCCGCCACCGGTGCGGTTGTGGCGGTG 118
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 835437 GCGCCGCGGTCAATACAGTCCGCTGCTGCGACGATCAAGCGCCCTTCTGCGCGGT 835496
OY 119 TTGCGCGCAGCTGTGTAAGGCTGTGCTGTGTCGATGCGCATCGCGCGGTGCGG 178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 835497 TTGCGCGCCCGCTGATGTGTAACGCGCCGACGCGGCGCGCGCGCGCGCGGT 835556
OY 179 TTCTGGCGGCTGTGATGCTTAATAATCGGTACGAC 216
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 835557 GGGCGCGCGATGTGTTGTCGCAACGCGCGCGCGCGG 835594
```

```
RESULT 12
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match          6.9%; Score 33.2; DB 4; Length 4411529;
Best Local Similarity 50.6%; Pred. No. 23;
Matches 80; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
```

```
OY 59 GCGCCAGAACTTCAGCCGCCAGAAATTGGCCGCCACCGGTGCGGTTGTGGCGGTG 118
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


QY 152 GTGTGCGATGGCCATGCGCGGTTCGTGGCGGTCGTGATTTGGCTCTAAATCGGTC 211
|| | ||| ||| ||| ||| | ||| | | ||| |||
Db 362 TTGACCTGGAGGCTCAGCGCAGTGGGTTTCAGGACCTTCTGTCCCGGTGATATGCTC 303
QY 212 AGAGCATGGACCAG 225
|| | |||
Db 302 AGGAGCGGCTCAG 289

Search completed: October 27, 2002, 20:15:39
Job time : 5939.31 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 14:35:03 ; Search time 122.627 Seconds
(without alignments)
6762.541 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483
Sequence: 1 atgcgtggttgctgcagg.....agtgattagcaccgaaaa 483

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

1: N.Genseq/03f802.*
2: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1980.DAT.*
3: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1981.DAT.*
4: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1982.DAT.*
5: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1983.DAT.*
6: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1984.DAT.*
7: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1985.DAT.*
8: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1986.DAT.*
9: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1987.DAT.*
10: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1988.DAT.*
11: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1989.DAT.*
12: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1990.DAT.*
13: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1991.DAT.*
14: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1992.DAT.*
15: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1993.DAT.*
16: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1994.DAT.*
17: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1995.DAT.*
18: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1996.DAT.*
19: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1997.DAT.*
20: /SDSI1/gcgdata/genseq/genseqn-emb1/NA1998.DAT.*
21: /SDSI1/gcgdata/genseq/genseqn-emb1/NA2000.DAT.*
22: /SDSI1/gcgdata/genseq/genseqn-emb1/NA2001A.DAT.*
23: /SDSI1/gcgdata/genseq/genseqn-emb1/NA2001B.DAT.*
24: /SDSI1/gcgdata/genseq/genseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	483	100.0	483	22	AAF86247	DNA sequence of E.
2	483	100.0	768	22	AAF86248	DNA sequence of C.
3	295	61.1	486	22	AAF86246	Ospa 17kD antigen
4	295	61.1	489	22	AAH79040	Piscirickettsia sa
5	118	24.4	486	22	AAF86254	PCR primer #4 used
6	108.4	22.4	110	22	AAF86252	PCR primer #2 used
7	102	21.1	102	22	AAF86255	PCR primer #5 used
8	92.4	19.1	94	22	AAF86253	PCR primer #3 used
9	74	15.3	111	22	AAF86251	PCR primer #1 used

C	10	72.4	15.0	110	22	AAE86256	PCR primer #6 used
C	11	39.4	8.2	2292	22	AAC45883	Arabidopsis thaliana
C	12	36.6	7.6	303	16	AAT18913	DNA encoding spide
C	13	36.6	7.6	303	16	AAT18912	DNA encoding spide
C	14	36	7.5	1305	17	AAT31799	Erythrobacter long
C	15	36	7.5	3251	23	ABL03846	Drosophila melanog
C	16	36	7.5	7791	22	ABL37335	Human musculuskele
C	17	35.6	7.4	292	21	AAA45539	Human secreted exp
C	18	35.6	7.4	2481	20	AAA90448	Human secreted pro
C	19	35.6	7.4	2481	22	AAS35277	Human cDNA encodin
C	20	35.6	7.4	2481	20	ABA69046	Human polynucleoti
C	21	35.4	7.3	571	20	AAAG61355	DNA encoding a hum
C	22	35.4	7.3	1290	19	AAV36619	Nucleotide sequenc
C	23	35.4	7.3	27541	22	AAD17185	Streptomyces nours
C	24	35.4	7.3	125401	22	AAD17186	Streptomyces nours
C	25	35.2	7.3	1479	21	AAAS1609	Hiv synthetic Gag
C	26	35	7.2	77536	21	AAA41651	Nucleotide sequenc
C	27	34.8	7.2	9407	23	AAAS59606	Propionibacterium
C	28	34.6	7.2	77536	21	AAA14651	Nucleotide sequenc
C	29	34.2	7.1	6273	23	ABL15731	Drosophila melanog
C	30	34.2	7.1	10855	23	ABL15730	Drosophila melanog
C	31	34	7.0	2652	20	AAK25386	Human Delta7-ster
C	32	34	7.0	2652	20	AAK25387	Human Delta7-ster
C	33	33.8	7.0	810	22	AAAX23387	Human Delta7-ster
C	34	33.8	7.0	933	22	AAAF66598	C glutamicum codin
C	35	33.8	7.0	9431	13	AAO20907	Corynebacterium gl
C	36	33.8	7.0	349980	22	AAH65529	DNA encoding the c
C	37	33.6	7.0	1164	22	AAH67598	C glutamicum codin
C	38	33.6	7.0	1479	21	AAA451625	C glutamicum codin
C	39	33.6	7.0	349980	22	AAH65532	C glutamicum codin
C	40	33.6	7.0	4403765	22	AAI196683	Mycobacterium tub
C	41	33.6	6.9	4411529	22	AAI196682	Mycobacterium tub
C	42	33.4	6.9	2888	17	AAT59268	Streptomyces prist
C	43	33.4	6.9	109519	22	AAAD08693	Micromonospora DNA
C	44	33.2	6.9	506	21	AACT5679	Human ORFX ORF1234
C	45	33.2	6.9	1047	22	AAAT7614	Nucleotide sequenc

ALIGNMENTS

	RESULT	1	
AAAF86247	ID	AAF86247 standard; DNA; 483 BP.	
AC	AC	AAF86247;	
AD	AD	11-JUL-2001 (first entry)	
DE	DE	DNA sequence of E. coli optimised ospA gene 17E2.	
KW	KW	Polyklohermich fish: Piscirickettsia salmonis; rickettsial pathogen;	
KW	KW	vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;	
SRS; 17E2; ds.	SRS; 17E2; ds.		
XX	XX		
OS	OS	Piscirickettsia salmonis.	
OS	OS	Synthetic.	
XX	XX		
FH	FH	location/Qualifiers	
FT	FT	1..483	
CDS	CDS	/tag= a	
FT	FT	/partial	
FT	FT	/product= "OspA"	
FT	FT	/note= "Genus specific 17kDa antigen, the sequence does	
XX	XX	not include a stop codon"	
PN	PN		
CA2281913-Al.	CA2281913-Al.		
PD	PD	17-MAR-2001.	
PF	PF	17-SEP-1999; 99CA-2281913.	
XX	XX		
17-SEP-1999;	17-SEP-1999;	99CA-2281913.	
XX	XX		

PN	CA2281913-A1.
XX	
PD	17-MAR-2001.
XX	
PF	17-SEP-1999;
XX	99CA-2281913
PR	17-SEP-1999;
	99CA-2281913


```

Db 286 ATGCGTGTTGCTGACGAGCAGCTCTGATCATATATCTCTGTTTCTGTCGGTTC 345
Qy 61 GCCAGAACTTCAAGCCCGCAGAGATGGGCGGCCACCGGTGGGTTGGCGGTTT 120
Db 346 GCCAGAACTTCAAGCCCGCAGAGATGGGCGGCCACCGGTGGGTTGGCGGTTT 405
Qy 121 GCCGCGCAGCTGTTGGTAAAGGCTGTGTCGTGTCATGAGGATGGGCGGTTT 180
Db 406 GCCGCGCAGCTGTTGGTAAAGGCTGTGTCGTGTCATGAGGATGGGCGGTTT 465
Qy 181 CTGGCGGCTGTGATTTGGCTCTAAATCGTCAGAGATGACAGCAGATAAATCAA 240
Db 466 CTGGCGGCTGTGATTTGGCTCTAAATCGTCAGAGATGACAGCAGATAAATCAA 525
Qy 241 CTGACCAAGTCTCTGAAAAAGTAAAGCCGCGCAGGTTACTGTTGGCGTAAATCCGAC 300
Db 526 CTGAACCAAGTCTCTGAAAAAGTAAAGCCGCGCAGGTTACTGTTGGCGTAAATCCGAC 585
Qy 301 ACCGCTAACAGCTACTGTGGAACCGGTTCCGACCTACGAGCTTACAAACAAAGGAA 360
Db 586 ACCGCTAACAGCTACTGTGGAACCGGTTCCGACCTACGAGCTTACAAACAAAGGAA 645
Qy 361 CGCGCTACAGATCTACTCCGGAATTTTCAGCAGAAAGCAGATGCGAGTCAAGAAACAG 420
Db 646 CGCGCTACAGATCTACTCCGGAATTTTCAGCAGAAAGCAGATGCGAGTCAAGAAACAG 705
Qy 421 GAAATCTACGCGACCGGCTGCTCCTCAGCCGAGTGGCGGTGGAGGATAGCAGCCGAA 480
Db 706 GAAATCTACGCGACCGGCTGCTCCTCAGCCGAGTGGCGGTGGAGGATAGCAGCCGAA 765
Qy 481 AAA 483
Db 766 AAA 768

RESULT 3
AAF86246
ID AAF86246 standard; DNA; 486 BP.
XX
AC AAF86246;
XX
DT 11-JUL-2001 (first entry)
XX
DE OspA 17kd antigen gene.
XX
KW Polkiothermic fish: piscirickettsia salmonis; rickettsial pathogen;
KM vaccine; OspA; salmonid rickettsial septicemia; rickettsial disease;
XX SRS; ds.
XX
OS Piscirickettsia salmonis.
XX
FH Key Location/Qualifiers
FT CDS 1..486
FT /tag- a
FT /partial
FT /product= "OspA"
FT /note= "Genus specific 17kda antigen, the sequence does
FT not include a stop codon"
XX
XX CA2281913-A1.
XX
PN 17-MAR-2001.
XX
PD 17-SEP-1999; 99CA-2281913.
XX
PF 17-SEP-1999; 99CA-2281913.
XX
PR 17-SEP-1999; 99CA-2281913.
XX
PA (KAYM/) KAY M W.
PA (BURI/) BURIAN J.
PA (KUZU/) KUZUK M A.
XX
XX Kay WW, Burian J, Kuzuk MA:
XX
XX

```

```

DR WPI: 2001-316844/34.
DR P-PsDB: AAB81126.
PT Method for protecting polkiothermic fish against salmonid rickettsial
PT septicemia and other rickettsial diseases comprises administering a
PT vaccine containing the OspA protein of Piscirickettsia salmonis.
XX
XX Disclosure; Fig 2B; 35pp; English.
XX
XX This invention relates to a method for the protection against infection
CC of a polkiothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC OspA in the form of a vaccine. The method is used for protecting animals,
CC particularly polkiothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
CC sequence represents DNA encoding the P. salmonis OspA protein. An OspA
CC protein with an N-terminal fusion partner is used in a vaccine to create
CC an anti-OspA antibody response.
XX
XX Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other;
SQ
Query Match 61.1%; Score 295; DB 22; Length 486;
Best Local Similarity 76.0%; Pred. No. 2, 9e-79;
Matches 364; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
Qy 5 GTGGTGCCTGACAGGAGCTCTCTGATCATATATCTGTTTCCGTGGGTTGCGCC 64
Db 8 GAGATGTTTTCAGAGTAGTAGTCTAATTTATACATGTTGTTTGTGGTGTGCC 67
Qy 65 AGAATCTACGCGCAGAGAGTTGGCGCCACCGGTGGGTTGGCGGTTTCCG 124
Db 68 AGAATCTTACGTCAGAGAGTGGACCTGCGAGCTGGGCTGTTGTTGGTCTG 127
Qy 125 GCCAGCTGTTGCTGTAAGGCTCTGCTGTGTCATGAGCCATGCGCGGTGCTCG 184
Db 128 GCCAGCTGTTGCTGTAAGGCTCTGCTGTGTCATGAGCCATGCGCGGTGCTCG 187
Qy 185 GCGGCTGATTTGGCTTAATTCGTCAGAGATGACAGCAGATAAATCAACTGA 244
Db 188 GTGATTAATTTGGTCTTAATTCGTCAGAGATGACAGCAGATAAATCAACTGA 247
Qy 245 ACCAGCTCTGGAAGAGTGAAGCCGCGCAGATGACTGTTGGCTTAATTCGACACCG 304
Db 248 ACCAGCTCTGGAAGAGTGAAGCCGCGCAGATGACTGTTGGCTTAATTCGACACCG 307
Qy 305 GTACAGCTACTCTGTGGAACCGGTTCCAGCTTACACAGGATGACAGGAGCC 364
Db 308 GCAATAGTTATAGTTGAGGCAAGTGTGCTTACACAGGATGACAGGAGGCTC 367
Qy 365 GTCAGAGTACTGCGGGAATTTTCAGCAGAAAGCCATGATGCGAGTCAAGAAACAGAGAA 424
Db 368 GCCAGCAATATTTGCGAGAAATTTTCAGCAGAAAGCCATGATGCGAGTCAAGAAACAGAGAA 427
Qy 425 TCTAGCGCAGCGGCTGCGCTCAGCGGATGGCGCTGGCAGATTAAGCAAGCAAGAA 483
Db 428 TTTACGGCAGCTGATGCGGCAACCGAGATGTTGTTGCAAGTCAATTCACAGAGAA 486

RESULT 4
AAH79040
ID AAH79040 standard; cDNA; 489 BP.
XX
AC AAH79040;
XX
DT 15-JAN-2002 (first entry)
XX
DE Piscirickettsia salmonis polynucleotide p10.6.
XX
XX Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
KM septicemia; SRS; surface antigen; vaccine; antibacterial; fish;
XX AMCC VR-1361; SS.
XX

```



```

AC AAF86252;
XX
XX 11-JUL-2001 (first entry)
DE PCR primer #2 used in cloning an optimisation of Ospa gene.
XX
XX Polkillothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KM vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
XX SRS; 17E2; fusion construct; PCR primer; ss.
XX
XX Piscirickettsia salmonis.
OS
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
PA (BUR/) BURIAN J.
PA (KUZ/) KUZIK M A.
XX
XX Kay WM, Burian J, Kuzik MA;
XX
XX WPI; 2001-316844/34.
XX
XX Method for protecting polkillothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of Piscirickettsia salmonis -
XX
XX Example 3; Fig 4B; 35pp; English.
XX
XX This invention relates to a method for the protection against infection
CC of a polkillothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkillothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the
CC invention.
XX
XX Sequence 110 BP; 13 A; 28 C; 45 G; 24 T; 0 other:
SO
Query Match 22.4%; Score 108.4; DB 22; Length 110;
Best Local Similarity 99.1%; Pred. No. 4.2e-23;
Matches 109; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 53 TGGGTTGGCCAGAACTTCAGCCGCGGAGAGTTGGCGGCGCCACCGGTGGGTTGG 112
DB 1 TGGGTTGGCCGCGCAAGATTCTCAGCCGCGGAGAGTTGGCGGCGCCACCGGTGG 60
OY 113 GCGGTGTGGCCGCGCAAGTTGGTTAAAGCTGTGCTGTGTCATG 162
DB 61 GCGGTGTGGCCGCGCAAGTTGGTTAAAGCTGTGCTGTGTCATG 110

```

RESULT 7

AAF86255/c
ID AAF86255 standard; DNA: 102 BP.

AC AAF86255;
XX

DT 11-JUL-2001 (first entry)

DE PCR primer #5 used in cloning an optimisation of Ospa gene.

KM Polkillothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KM vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
KM

```

KM SRS; 17E2; fusion construct; PCR primer; ss.
XX
XX Piscirickettsia salmonis.
OS
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
PA (BUR/) BURIAN J.
PA (KUZ/) KUZIK M A.
XX
XX Kay WM, Burian J, Kuzik MA;
XX
XX WPI; 2001-316844/34.
XX
XX Method for protecting polkillothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of Piscirickettsia salmonis -
XX
XX Example 3; Fig 4B; 35pp; English.
XX
XX This invention relates to a method for the protection against infection
CC of a polkillothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkillothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the
CC invention.
XX
XX Sequence 102 BP; 14 A; 23 C; 30 G; 35 T; 0 other:
SO
Query Match 21.1%; Score 102; DB 22; Length 102;
Best Local Similarity 100.0%; Pred. No. 3.6e-21;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 334 ACCTACGAGCGTTACACAAACAGAGAGCGCGTCAGAGTACTGCCGGAATTCAGCAG 393
DB 102 ACCTACGAGCGTTACACAAACAGAGAGCGCGTCAGAGTACTGCCGGAATTCAGCAG 43
OY 394 AAGCCATGATCCAGAGTCAAGAAACAGAAATCTACGGCACC 435
DB 42 AAGCCATGATCCAGAGTCAAGAAACAGAAATCTACGGCACC 1

```

RESULT 8
AAF86253
ID AAF86253 standard; DNA: 94 BP.

AC AAF86253;
XX

DT 11-JUL-2001 (first entry)

DE PCR primer #3 used in cloning an optimisation of Ospa gene.

KM Polkillothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KM vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
KM SRS; 17E2; fusion construct; PCR primer; ss.

OS Piscirickettsia salmonis.

XX CA2281913-A1.

XX 17-MAR-2001.

```

PF 17-SEP-1999; 99CA-2281913.
XX
PR 17-SEP-1999; 99CA-2281913.
XX
PA (KAYW/) KAY W W.
XX
PA (BURI/) BURIAN J.
XX
PA (KUZV/) KUZYK M A.
XX
PI Kay WW, Burian J, Kuzyk MA;
XX
DR WPI: 2001-316844/34.
XX
PT Method for protecting poikilothermic fish against salmonid rickettsial
XX septicemia and other rickettsial diseases comprises administering a
XX vaccine containing the Ospa protein of Piscirickettsia salmonis
XX
PS Example 3; Fig 4B; 35pp: English.
XX
CC This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly poikilothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the
CC invention.
XX
SQ Sequence 94 BP; 18 A; 20 C; 35 G; 21 T; 0 other;
XX
Query Match 19.1%; Score 92.4; DB 22; Length 94;
Best Local Similarity 98.9%; Pred. No. 2.7e-18;
Matches 93; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 139 AAGGCTGTGCTGTGTCGATGCGCATGCGGCGTGGCTGTGCGGTGATGGC 198
DB 1 AAGGCTGTGCTGTGTCGATGCGCATGCGGCGTGGCTGTGCGGTGATGGC 198
QY 199 TCTAATCGGTGACGACATGACGACGACGATGATA 232
DB 61 TCTAATCGGTGACGACATGACGACGACGATGATA 94
XX
RESULT 9
AAF86251
ID AAF86251 standard; DNA; 111 BP.
XX
AC AAF86251;
XX
DT 11-JUL-2001 (first entry)
XX
DE PCR primer #1 used in cloning an optimisation of Ospa gene.
XX
KW Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;
KW SRS; 17E2; fusion construct; PCR primer; ss.
XX
OS Piscirickettsia salmonis.
XX
CA2281913-Al.
XX
PN 17-MAR-2001.
XX
PD 17-SEP-1999; 99CA-2281913.
XX
PE 17-SEP-1999; 99CA-2281913.
XX
PR 17-SEP-1999; 99CA-2281913.
XX
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZV/) KUZYK M A.
XX

```

```

PI Kay WW, Burian J, Kuzyk MA;
XX
DR WPI: 2001-316844/34.
XX
PT Method for protecting poikilothermic fish against salmonid rickettsial
XX septicemia and other rickettsial diseases comprises administering a
XX vaccine containing the Ospa protein of Piscirickettsia salmonis
XX
PS Example 3; Fig 4B; 35pp: English.
XX
CC This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly poikilothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the
CC invention.
XX
SQ Sequence 111 BP; 16 A; 32 C; 30 G; 33 T; 0 other;
XX
Query Match 15.3%; Score 74; DB 22; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.1e-12;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 ATGCGGTGTTGCCGACGAGGACGCTCTGATCATTTATCTGTTCTGCGGGGTGC 60
DB 38 ATGCGGTGTTGCCGACGAGGACGCTCTGATCATTTATCTGTTCTGCGGGGTGC 60
QY 61 GCCCAGAACTTCAG 74
DB 98 GCCCAGAACTTCAG 111
XX
RESULT 10
AAF86256/C
ID AAF86256 standard; DNA; 110 BP.
XX
AC AAF86256;
XX
DT 11-JUL-2001 (first entry)
XX
DE PCR primer #6 used in cloning an optimisation of Ospa gene.
XX
KW Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;
KW SRS; 17E2; fusion construct; PCR primer; ss.
XX
OS Piscirickettsia salmonis.
XX
CA2281913-Al.
XX
PN 17-MAR-2001.
XX
PD 17-SEP-1999; 99CA-2281913.
XX
PE 17-SEP-1999; 99CA-2281913.
XX
PR 17-SEP-1999; 99CA-2281913.
XX
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZV/) KUZYK M A.
XX
PI Kay WW, Burian J, Kuzyk MA;
XX
DR WPI: 2001-316844/34.
XX
PT Method for protecting poikilothermic fish against salmonid rickettsial
XX septicemia and other rickettsial diseases comprises administering a
XX vaccine containing the Ospa protein of Piscirickettsia salmonis
XX

```

PS Example 3; Fig 4B; 35pp; English.
XX
CC This invention relates to a method for the protection against infection
CC of a polkiethermic fish by the bacterial pathogen, *Piscirickettsia*
CC salmonis. The method comprises administering an immunogenic amount of a
CC *P. salmonis* specific antigen termed OspA, or an immunogenic fragment of
CC OspA in the form of a vaccine. The method is used for protecting animals,
CC particularly polkiethermic fish, against the bacterial pathogen
CC *P. salmonis*. The method is also useful for protecting against salmonid
CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the *P. salmonis* OspA gene. The OspA gene is used in the method of the
CC invention.
XX
SO Sequence 110 BP; 17 A; 34 C; 27 G; 32 T; 0 other;

Query Match 15.08; Score 72.4; DB 22; Length 110;
Best Local Similarity: 98.6%; Pred. No. 3.3e-12;
Matches 73; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 410 GTCGAAACGGAATCTACGCGACCGCGCTCCCTCAGCCGAGTGGCCGCGCAGGTGA 469
Db 110 GTCGAAACGGAATCTACGCGACCGCGCTCCCTCAGCCGAGTGGCCGCGCAGGTGA 51
Qy 470 TTACACCGGAAAA 483
Db 50 TTACACCGGAAAA 37

RESULT 11
AAC45883/C
ID AAC45883 standard; DNA: 2292 BP.
XX
AC AAC45883;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 48114.
XX
DE Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS
XX Arabidopsis thaliana.
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

```

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

```

```

PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 8.2%; Score 39.4; DB 21; Length 2292;
 Best Local Similarity 59.3%; Pred. No. 0.11;
 Matches 67; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```

QY 85 GTTGGCGCGCCACCGGTGCGGTGTTGGGGGTGTTGCCGCGCAGCTGTCGTAAGGC 144
DB 1256 GTTGGAGGTGTCCTCCGGTGTGTGTTGGAGAGTCTCGTGTGTTGGAGAGTCTC 1197
QY 145 TCTGTCGTGTGTCGATGCGCATCGCGGTGCGTCTGCGCGTCTGATTTGG 197
DB 1196 TCTGTGTGTGTGTTGGAGAGTTCCTCCGGTGTGTGTTGGAGGTGTTCTGG 1144

```

RESULT 12

ID AAT18913 standard; DNA; 303 BP.
 XX AAT18913;

DE 17-JAN-1997 (first entry)

DE DNA encoding spider dragline variant, DP-1B.9 monomer.

KW Spider: dragline protein; variant; monomer; polymer;
 KW fibre forming region; Spidroin 1; Nephrin clavipect; DPL; mimic;
 KW DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
 KW rope; surgical suture; implant; reinforcement; film; coating; ss.
 OS Synthetic.

FN WO9429450-A2.

XX 22-DEC-1994.

PF 15-JUN-1994; 94MO-US06689.

PR 15-JUN-1993; 93US-0077600.

PA (DUPO) DU PONT DE NEMOURS & CO E. I.

PI Fahnestock SR;

DR WPI: 1995-036479/05.

XX P-PSDB; AAR99054.

PT New synthetic variants of spider dragline protein - for making
 PT fibres useful as clothing, surgical silk, plastic reinforcement
 PT etc.; also related DNA, vectors and transformed cells
 XX
 PS Claim 5; Page 122; 168pp; English.

CC This sequence encodes a synthetic spider dragline variant monomer,
 CC DP-1B.9. The sequence of the DP-1B.9 polymer is given in AAR99055.
 CC The polypeptide monomer is a variant based on a consensus sequence
 CC derived from the fibre forming regions of spider dragline protein,
 CC esp. the natural protein 1 (Spidroin 1) from *Nephila clavipes*. This
 CC DNA sequence may be used in the recombinant production of the
 CC variant protein in a recombinant host, e.g. *E. coli* or *Bacillus*
 CC *subtilis*. Synthetic analogues of DPL were designed to mimic the
 CC repeating consensus sequence of the natural protein and the pattern

CC of variation among individual repeats. DP-1B analogues were designed
CC by reordering the four repeats within the monomer of DP1A. This
CC monomer exhibits all of the regularities of (1)-(5) below. In addition,
CC it exhibits a regularity of the natural sequence which is not shared by
CC DP-1A, namely that a repeat in which both GYG and GGG are deleted is
CC generally preceded by a repeat lacking the entire poly-alanine repeat,
CC with one intervening repeat. The sequence of DP-1B matches the natural
CC sequence more closely over a more extended segment than does DP-1A. The
CC individual repeats differ from the consensus according to the pattern:
CC (1) the poly-alanine sequence varies in length from 0-7
CC residues; (2) when the entire poly-alanine sequence is deleted,
CC so also is the surrounding sequence encompassing AGRGGLGGGAGANGG;
CC (3) aside from the poly-alanine sequence, deletions usually
CC encompass integral multiples of three consecutive residues;
CC (4) deletion of GYG is generally accompanied by deletion of GRG
CC in the same sequence; and
CC (5) a repeat in which the entire poly-alanine sequence is
CC deleted is generally preceded by a repeat containing six alanine
CC residues.
CC The proteins may be used to produce fibres of high tensile strength and
CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials
CC for implants, plastic reinforcements, films, coatings, etc.
CC
SQ Sequence 303 BP; 43 A; 69 C; 138 G; 53 T; 0 other;

Query Match 7.6%; Score 36.6; DB 16; Length 303;
Best Local Similarity 57.4%; Pred. No. 0.34;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 85 GTTGGCCGCGCACCCGCTGCTGGGCGGCTTCCGCCAGCTTTCGTAAGGC 144
DB 87 GTGCGGACAGCTGCTGGTGGTACAGCGGCTTCTGCACAAAGGCGCGTCAAGGC 146
QY 145 TCTGCTGCTGTGTCGATGCCATCGCGGCTTCTGCGCGCTGATTTGCT 199
DB 147 GCTGTGTCACAGCAGCTGCCGCTGCGGTGACAGCCAGGTGATGTGGCT 201

RESULT 13
AAT18912
ID AAT18912 standard; DNA: 303 BP.
XX
AC AAT18912;
XX
DT 15-JAN-1997 (first entry)
XX
DE DNA encoding spider dragline variant, DP-1A.9 monomer.
XX
KM Spider; dragline protein; variant; monomer; polymer;
KM fibre forming region; Spidroin 1; Nephila clavipes; DPI; mimic;
KM DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
KM rope; surgical suture; implant; reinforcement; film; coating; ss.
XX
OS Synthetic.
XX
PN MO9429450-A2.
XX
PD 22-DEC-1994.
XX
PF 15-JUN-1994; 94MO-US06689.
XX
PR 15-JUN-1993; 93US-0077600.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Farnesstock SR;
XX
XX WPI: 1995-036479/05.
DR P-PSDB; AAR99052.
XX
XX New synthetic variants of spider dragline protein - for making
PT fibres useful as clothing, surgical silk, plastic reinforcement
PT etc., also related DNA, vectors and transformed cells

XX
PS Claim 2; Page 121; 168pp; English.
XX
CC This sequence encodes a synthetic spider dragline variant monomer,
CC DP-1A.9. The sequence of the DP-1A.9 polymer is given in AAR99053.
CC The polypeptide monomer is a variant based on a consensus sequence
CC derived from the fibre forming regions of spider dragline protein,
CC esp. the natural protein 1 (Spidroin 1) from Nephila clavipes. This
CC DNA sequence may be used in the recombinant production of the
CC variant protein in a recombinant host, e.g. E. coli or Bacillus
CC subtilis. Synthetic analogues of DPI were designed to mimic the
CC repeating consensus sequence of the natural protein and the pattern
CC of variation among individual repeats. DP-1A analogues are composed
CC of a tandemly repeated 101 amino acid monomer which comprises four
CC repeats which differ from the consensus sequence given in AAM06201,
CC according to the pattern (1)-(5) given below. This 101 amino acid
CC monomer is repeated 1-16 times in a series of analogue proteins. The
CC individual repeats differ from the consensus according to the pattern:
CC (1) the poly-alanine sequence varies in length from 0-7
CC residues; (2) when the entire poly-alanine sequence is deleted,
CC so also is the surrounding sequence encompassing AGRGGLGGGAGANGG;
CC (3) aside from the poly-alanine sequence, deletions usually
CC encompass integral multiples of three consecutive residues;
CC (4) deletion of GYG is generally accompanied by deletion of GRG
CC in the same sequence; and
CC (5) a repeat in which the entire poly-alanine sequence is
CC deleted is generally preceded by a repeat containing six alanine
CC residues.
CC The proteins may be used to produce fibres of high tensile strength and
CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials
CC for implants, plastic reinforcements, films, coatings, etc.
CC
SQ Sequence 303 BP; 43 A; 69 C; 138 G; 53 T; 0 other;

Query Match 7.6%; Score 36.6; DB 16; Length 303;
Best Local Similarity 57.4%; Pred. No. 0.34;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 85 GTTGGCCGCGCACCCGCTGCTGGGCGGCTTCCGCCAGCTTTCGTAAGGC 144
DB 138 GCTGCGGACAGCTGCTGGTGGTACAGCGGCTTCTGCACAAAGGCGCGTCAAGGC 197
QY 145 TCTGCTGCTGTGTCGATGCCATCGCGGCTTCTGCGCGCTGATTTGCT 199
DB 198 GCTGTGTCACAGCAGCTGCCGCTGCGGTGACAGCCAGGTGATGTGGCT 252

RESULT 14
AAT31799/c
ID AAT31799 standard; DNA: 1305 BP.
XX
AC AAT31799;
XX
DT 20-SEP-1996 (first entry)
XX
DE Erythrobacter longus Lycopene cyclase gene.
XX
KM Phytoene dehydrogenase; Erythrobacter longus; Erwinia hedicola; probe;
KM open reading frame; lycopene cyclase; Rhodobacter sphaeroides;
KM beta-carotene; food colourant; additive; anti-cancer; ds.
XX
OS Erythrobacter longus.
XX
PN JF08089241-A.
XX
PD 09-APR-1996.
XX
PF 30-SEP-1994; 94JP-0236621.
XX
PR 30-SEP-1994; 94JP-0236621.
XX
PA (ASAH) ASAH KASEI KOGYO KK.
XX

DR WPI: 1996-23337/24.
 DR P-PSDB; AAR95698.
 XX Erythrobacter sp. phytoene dehydrogenase and lycopene cyclase genes
 PT - for the produ. of beta-carotene useful as a food colourant
 XX
 PS Example 1; Page 6-7; 7pp; Japanese.
 CC This is the nucleotide sequence encoding the lycopene cyclase enzyme
 CC from Erythrobacter longus ATCC 14126. The gene was isolated from an
 CC E. longus genomic DNA library using a 1 kb fragment of the phytoene
 CC dehydrogenase gene from Erwinia hebelicola as a probe. The probe isolated
 CC a nucleotide fragment which contained two open reading frames (ORF).
 CC ORF2 (1578 bp) encodes the phytoene dehydrogenase (AAT31798) whereas ORF1
 CC (1302 bp) encodes lycopene cyclase. The two ORFs were inserted into the
 CC plasmid pRK415 to generate pRKCD. This plasmid was transformed into
 CC Rhodospirillum rubrum cells for production of the proteins. The phytoene
 CC dehydrogenase and lycopene cyclase can be used to produce beta-carotene
 CC for use as a food colourant or as a food additive with anti-cancer
 CC effects.
 XX
 XX Sequence 1305 BP; 251 A; 392 C; 403 G; 259 T; 0 other;
 S0
 Query Match 7.5%; Score 36; DB 17; Length 1305;
 Best Local Similarity 56.9%; Pred. No. 0.95;
 Matches 66; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 QY 115 GGTGTTCCGCGCAGCTGTTCGTAAGCTCTGCTGCTGTCGATGCGCCATCGCGCGT 174
 Db 900 GGTTAGCTGTGTCAGCGGATGGTAAACCCGCCGCCGCCGCAATGCGCAACCCCGGG 841
 QY 175 GCGGTTCTGGGGGCTGTGATTTGCTTAATTCGTCAGAGCAGTACGACGACGAGA 230
 Db 840 AATGCCACTTCTGCTGATAGCGGCAAAATCGCCGCCGCCGTCAGACGACGCGCAAGA 785
 RESULT 15
 ABL03846
 ID ABL03846 standard; cDNA; 3251 BP.
 XX
 AC ABL03846;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 6020.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-061415O.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR P-PSDB; ABB59743.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 6020; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB1616176-AB1616175), expressed DNA
 CC sequences (AB1616176-AB1616175) and the encoded proteins
 CC (AB57737-AB57737).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 S0 Sequence 3251 BP; 859 A; 752 C; 758 G; 882 T; 0 other;

Query Match 7.5%; Score 36; DB 23; Length 3251;
 Best Local Similarity 46.4%; Pred. No. 1.4;
 Matches 117; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
 QY 52 GTGGTTGCGCCGACGACTTCAAGCGCGGAGATTGGCGCGCCACCGGTGGTTCG 111
 Db 1279 GTGATGATGTCACGACGACGACGACGACGACGACGACGACGACGACGACGAT 1338
 QY 112 GCGGTTGCGCGCCGACGCTGTTGCTTAAGGCTGTGCTGTGTCGATGCGCATCGGC 171
 Db 1339 GTGCGGCTTCTTGGCGCAGCTCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGG 1398
 QY 172 GTGCGGTTCTGCGCGGCTGATTTGCTCTAAATCGTCAGAGCATGAGCAGCAGAT 231
 Db 1399 GCGGAGGCTGGCGGACACGCTCTTGAAGATGTTGCTGACGACGACGACGAGGAG 1458
 QY 232 AAATCAACTGACACGATCTCTGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 291
 Db 1459 GTGACCTCTTGAAGCTCTGCAATTTGGAAGGAGGAGGAGGAGGAGGAGGAGGAG 1518
 QY 292 AATCGGACACG 303
 Db 1519 CTGCGGCGGACG 1530

Search completed: October 27, 2002, 16:40:46
 Job time : 129.627 secs

Query Match	61.1%; Score 295; DB 6; Length 489;
Best Local Similarity	76.0%; Pred. No. 1.9e-61;
Matches 364; Conservative	0; Mismatches 115; Indels 0; Gaps 0;
OY	5 GTGGTTGCTGAGGAGGAGCTCTGATCATTTATCTGCTTTCTGCTGTTGCGGCGCC 64
Db	8 GAGGATGTTTGAAGGTACTAGCTAATATATATATATATATATATATATATATATAT 67
OY	65 AGAATCTTACGCCGAGAGTGGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTG 124
Db	68 AGAATCTTATGTTGCTAAGAGAGTGGAGCTGGGAGCTGGGAGCTGGGAGCTGGG 127
OY	125 GCCAGCTGTTGCTGTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 184
Db	128 GCCAGCTGTTGCTGTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 187
OY	165 GCGGTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 244
Db	168 GTGGATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 247
OY	245 ACCAGCTCTGTAAGAGTGAAGCGGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTG 304
Db	248 ACCAGAGTTTGAAGAGTGAAGCGGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTG 307
OY	305 GTACAGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 364
Db	308 GCAATGTTATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 367
OY	365 GTGCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 424
Db	368 GCGGCAATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 427
OY	425 TCTACGCGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 483
Db	428 TTTACGGACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 486
RESULT 2	
AF184152	4983 bp DNA linear BCT 01-0CT-2000
LOCUS	Piscirickettsia salmonis alanine racemase (alr) gene, partial cds;
DEFINITION	and BAX (bax), DNA repair enzyme Rada (rada), 17 kDa antigen
ACCESSION	(ospa), and transposase (tnpa) genes, complete cds.
VERSION	AF184152.1 GI:10441343
KEYWORDS	
SOURCE	Piscirickettsia salmonis.
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group; Piscirickettsia.
REFERENCE	1 (bases 1 to 4983)
AUTHORS	Kuzuy, M.A., Burian, J., Thornton, J.C. and Kay, W.W.
TITLE	Identification of a genus-common Rickettsial surface antigen in the salmonid pathogen Piscirickettsia salmonis
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 4983)
AUTHORS	Kuzuy, M.A. and Kay, W.W.
TITLE	Direct Submission
JOURNAL	Submitted (10-SEP-1999) Biochemistry & Microbiology, University of Victoria, Petch Building, PO Box 3055, Victoria, BC V8W 3P6, Canada
FEATURES	Location/Qualifiers
source	1..4983
	/organism="Piscirickettsia salmonis"
	/strain="LF-89"
	/db_xref="taxon:1238"
	/clone="PB12"
gene	<1..501
	/gene="alr"
CDS	<1..501
	/gene="alr"
	/codon_start=1
	/transl_table=1
	/evidence=not_experimental
	/product="alanine racemase"
	/protein_id="AAG17001.1"
	/db_xref="GI:10441347"
	/translation="NSAIFNFSEYERYDWRPCLIMLYGISPEADKNGVDLEQPYMHV
	VSRILSVKOLROGESVYGATGWCCEPMOQGLISLGGYGPPLASGTPPIVROGR
	ALIGRVSMOIALIDLRCRDPACGAEVYWGDDLPPEETARHAGTTATETLCVMPLRA
	PIYIQW"
	complement(526..1257)
	/gene="bax"
	complement(526..1257)
	/gene="bax"
	/codon_start=1
	/transl_table=11
	/evidence=not_experimental
	/product="BAX"
	/protein_id="AAG17002.1"
	/db_xref="GI:10441348"
	/translation="MSLAVITTYAVLAQPLPSAPNSKTQINTAVSODIQSIKOEFTI
	KIQSNTINQIILQKROHILSLQKSKSKOBTLOQLAHQYKLNLIHFHHPDFS
	LLIKRVDIIPNALISAOAINESNMGSRFAVEGNFEGMCHYPCGGLIPKRPANNH
	WEVANYSSMTASVRAVYIHPLNTHNAYQALRLDIRAHIRANHOVSFAFKLAEGFLAVSIK
	GTKYVELIOTIITRYLSKESTQIP"
	1373..2743
	/gene="rada"
	/evidence=experimental
	1373..2743
	/gene="rada"
	/codon_start=1
	/transl_table=11
	/product="DNA repair enzyme Rada"
	/protein_id="AAG16998.1"
	/db_xref="GI:10441344"
	/translation="MSKQNOYVSCDGGIAARKWLQCPHCQOMSLSEVKEVLPNRP
	GSQRPAGPAGIDAPRKCLSDITEQISROQVGIQEDPRVJGGGLVGAVALIGGDP
	GICKSTLLQTSVNCIOFGKVLVYTGESLEQVTLRSKRLGSDVDRLAEQOVER
	ILKAAEIEQPKVLIVDSIQITPESLOSAPGGVAQVRRSAATLTOFAKRTGCTCLFNG
	HYTKGALAGPRLVLEHMDTVLYFEGEDSRRRLRAVKNRGANEGITRAMETGL
	KTVNSPSAIFLSRYEDLOPGSVYVAMESTRPLIVEVALVDESHSSNPRRTAVGLDQ
	ORLMLAVLNKHGGIATSDQVFTINVGMRITETADLALLACVSSLKGLASOE
	LIVFEIIGLSGEIRVQRLERIKBAVNHGFTQALVPLANCPKQKVDGIEIVGVKHE
	DALAVL"
	2834..3322
	/gene="ospa"
	/evidence=experimental
	2834..3322
	/gene="ospa"
	/note="similar to the rickettsial 17 kDa antigen"
	/codon_start=1
	/transl_table=11
	/product="17 kDa antigen"
	/protein_id="AAG17000.1"
	/db_xref="GI:10441346"
	/translation="NNRCGLQSSLLIISVFLVCAQNFNSROEYGAATGAVGCVAGQ
	LFNGSGRVAMALIGAVLGLISKTIQSGMOQDKIKLNSGLEKVGKAGOVVRWNPDT
	GNISYVEPVRTYQRYNKOERQYCRERQKAMLAGQKQELIYTAGCRPDORNOVIST
	EX"
	3560..4849
	/organism="Piscirickettsia salmonis"
	/strain="LF-89"
	/db_xref="taxon:1238"
	/clone="PB12"
	/insertion_seq="insertion sequence-like element"
	3560..3846
	complement(3842..4561)
	/gene="tnpa"
	complement(3842..4561)
	/gene="tnpa"
	/codon_start=1
	/transl_table=11
	/product="transposase"
	/protein_id="AAG16999.1"
	/db_xref="GI:10441345"

/translation="MDTLTSLFCVIDDFCOELLPQMAILLDETNRKRNKPSOMSTS
EIMTIMYFHKSNRYRNFKMYLHVIVKSGMYKYPNSVSNRWFELMPSLILPFCFLA
AOGKTAATYFVNDSTTLRCHKERASONRAFKLAKSKSTMGWYGFKLITIVNMG
ELMAFKMSKATTDORVLPKMAENLTGK11IGDKYISQKLPDLYERGLDLYIKRKN
MKNKLVIMDKILTKRGI1A"

repeat_region 1362 a 778 c 1265 g 1578 t

BASE COUNT

Query Match 61.18; Score 295; DB 1: Length 4983;
Best Local Similarity 76.08; Pred. No. 2e-61;
Matches 364; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

5 GTGGTTCGCGCAGGAGCGCTGATCATTCCTCTGTTTCCGTGGGTGGCCGC 64
Db 2841 GAGCATTTTTCAGAGTAGTACTAATTAATCAAGTGGTTTGTGGTGGCC 2900
QY 65 AGAATTACACCCGACAGAGTTGGCGCGCACCGGTGGGTGGCGGTGGTCCG 124
Db 2901 AGAATTAGTCGTCAGAGAGTGGAGCTGGAGCTGGGGGCTGTGGGGGTGCTG 2960
QY 125 GCCAGCTGTCGTAAGGCTGTGCTGATGATGCCATCGCGGTGGCTTGG 184
Db 2961 GCCAGCTGTCGTAAGGCTGTGCTGATGATGCCATCGCGGTGGCTTGG 3020
QY 185 GCGGTGATTTGGCTTAAATGCGTCAGAGCATGACGACGATTAATCAACTGA 244
Db 3021 GTGGATTAATTTGGCTTAAATGCGTCATGATGATGATGACGACGATTAATCAACTGA 3080
QY 245 ACCAGTCTGTAAGAAAGTGAAGCCGCGCAGTGTACTGCTGGCTAATCCGACACG 304
Db 3081 ACCAGTCTGTAAGAAAGTGAAGCCGCGCAGTGTACTGCTGGCTAATCCGACACG 3140
QY 305 GTAACAGCTACTGTGTAAGACGGTGGCACCTAACCAGCTTACACAAAGACGCC 364
Db 3141 GCATATGATTAATGTTGAGGACGACGTAATCTTACACGCTTACAAAGACGAGCTC 3200
QY 365 GTGCGAGTCTGCGGCAATTCAGCAAGAACGATGATGCGAGTCAAGAAAGGAA 424
Db 3201 GCCAGCAATTAATGTCAGAAATTTACGAAAGGCGATGATGCGAGGCAAGAAAGCA 3260
QY 425 TCTACGCGACCGCGCTGCGCTCAGCGGATGCGCGTGGAGTATAGCACCGAAAA 483
Db 3261 TTTACGCGACCTGCAATGCGGCAACCGATGCTGTGGCAAGTATTCAACACAAAA 3319

RESULT 3
RIR17KGA 537 bp DNA linear BCT 04-FEB-1999
LOCUS Rickettsia japonica 17K genus-common antigen gene, complete cds.
DEFINITION D16515.1 GI:695412
ACCESSION D16515.1 GI:695412
VERSION 17K genus-common antigen; 17K dalton protein.
KEYWORDS Rickettsia japonica (sub-species: yH) DNA.
SOURCE Rickettsia japonica
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsiidae; Rickettsia; spotted fever group.
1 (bases 1 to 537)
Yoshida, Y.

REFERENCE
AUTHORS Direct Submission
TITLE Submitted (05-JUL-1993) Yoshiya Yoshida, Kanagawa Prefectural
JOURNAL Pabulic Health Laboratory; 52-2 Nakao-cho, Asahi-ku, Yokohama,
Kanagawa 241, Japan (Tel:045-363-1030, Fax:045-363-1037)
2 (bases 1 to 537)
REFERENCE
AUTHORS Furuya, Y., Katayama, T., Yoshida, Y. and Kaho, I.
TITLE Specific amplification of Rickettsia japonica DNA from clinical
JOURNAL specimens by PCR
MEDLINE J Clin. Microbiol. 33 (2), 487-489 (1995)
FEATURES
source Location/Qualifiers
1. 537
/organism="Rickettsia japonica"
/sub-species="yH"

/db_xref="taxon:35790"
/note="propagated in Vero cells"
58.537
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="BA03965.1"
/db_xref="GI:1435056"

CDS

BASE COUNT 176 a 92 c 129 g 140 t

Query Match 9.68; Score 46.6; DB 1: Length 537;
Best Local Similarity 47.28; Pred. No. 0.5;
Matches 176; Conservative 0; Mismatches 194; Indels 3; Gaps 1;

QY 100 GGTGCGGTGGGCGGTGGTGGCGGCGAGCTGTGCTAAGAGCTGCTGCTGCTG 159
Db 163 GCGGTGCTGGTGGCGGCTTACTTCTGCTCAATTCGTAAGGCGACAGCACTGTG 222
QY 160 ATGGCCATCGCGGTGGGTGGGCGGCTGATGCTTAAATCGTCAAGCATG 219
Db 223 GAGTAGTGTGTAGTGCATTAATCTTGAGAGAGTCTTGTGGTGAACAATCGGTGATG 282
QY 220 GACCAAGCATTAATCAAACTGACCACTGCTTGGAAAAAGTGAAGCCGCGCAGTT 279
Db 283 GATGAGCAGATGAAACATCTTGACAGCTTACCTCACAGAGAGCTTTA--GAACAGCT 339
QY 280 ACTGTTGGGCTAATCGGACACCGCTACAGCTACTCTGNSAACGGTTGCGACCTAC 339
Db 340 CTTAGTGTGTAGTACGTGAATGCGCTTAATCCGATTAACGCAATTAACGTTACTAACA 399
QY 340 CAGCGTTACAAACAAGAACCGCTGACGAGTACTGCGGATTCACAGCAAAAGCC 399
Db 400 CTTAATAAATCTTAGAATAGCATGTCATATATTCGCTGAGTACATCAAAAGCTT 459
QY 400 ATGATCGAGTCAAGAAACAGAAATCTACGCGCACCGGTGCTCAGACCGGATGCGCC 459
Db 460 GTAAATAGCGGAAACAACAACAAGCATACGTAATGCTGCGCCCAACCTGACGAGCA 519
QY 460 TGGCAGGATTA 472
Db 520 TGGCAAGTTGTA 532

RESULT 4
RPXX04 237523 bp DNA linear BCT 11-NOV-1998
LOCUS Rickettsia prowazekii strain Madrid E, complete genome; segment
DEFINITION 4/4.
ACCESSION AJ235273 AJ235269
VERSION AJ235273.1 GI:3861237
KEYWORDS complete genome.
SOURCE Rickettsia prowazekii.
ORGANISM Rickettsia prowazekii.
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsiidae; Rickettsia; typhus group.
1 (bases 1 to 237523)
Andersson, S.G., Zomrodipour, A., Andersson, J.O.,
Sicheritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K.,
Eriksson, A.S., Winkler, H.H. and Kurland, C.G.
The genome sequence of Rickettsia prowazekii and the origin of
mitochondria
Nature 396 (6707), 133-140 (1998)

REFERENCE
AUTHORS Andersson, S.G.E.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-1998) S.G.E. Andersson,
siv.andersson@molbio.uu.se, Dept. of Molecular Biology, University
of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN

FEATURES
source
location/Qualifiers
1. .237523
/organism="Rickettsia prowazekii"
/strain="Madrid E"
/db_xref="taxon:782"
164. .1123
/gene="Rp701"
164. .1123
/gene="Rp701"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA15137.1"
/db_xref="GI:3861238"
/translation="MAEPIITLLFGKRDILRIRNOKALNKLIHTPILOVNESHS
DKERISEILEIKSYKECKIPITFINIQSNNNAVPLFRADODILNQNLIKTTITFDKN
ALPONSLSKYSWKQIIEKVDHVEFTENADNLSIADIIPKDKTTITDILRLVSYFN
NLIYNHTNHSGLTIYKVKDLIKAKNODSKIEITETPISYDEANILITAFGLITS
EDQIYGLKEIHEILKDPNNAENFKKYSQISQFCKDKDKDKNIDSNFTQKVI
NDSNKDIQVEHITSYVODKEKQSPQGLFRIRNFYKNIYKQ"
2074. .3348
/gene="Rp702"
2074. .3348
/gene="Rp702"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA15138.1"
/db_xref="GI:3861239"
/translation="MTTIVIVIIIMISLSALPAETATATASSPKRIKLKTAGNKR
AKTVLELVKKKREKRVIGLLIGNSLINICTIATTLTISLGNNGTIVASVAFII
VFAEVVPKAIYAKPEOLITKMSSTIYFLKRPKNIVTADYITKICFPFRINLQ
ISGEVEYRGVIEHHOEGGVYKSENNLGGIILIRNMTYSEIMTHRSNIALINDIPY
ELIKTLTSAHRIRIPLOMDNRNIIIGLTKULLKALFENNDKKVDINTLTPPM
FITDNALVDQLHAFRRNRHPCVVDVETGLITLEDVIEIVGPIIDEDRLN
EIIQESNTEFLIKGTTTIRDIRNEDLNMISDEDANTAGLIHKIARIPOGVIKIF
NFKIILKKIANKIDSVKITVLPMTETITSE"
3341. .5353
/gene="Rp703"
3341. .5353
/gene="Rp703"
/codon_start=1
/transl_table=11
/product="CYTOCHROME C-TYPE BIOGENESIS PROTEIN CCMF
(comp)"
/protein_id="CA15139.1"
/db_xref="GI:3861240"
/translation="MSKIGNFLLLTMCISGLITLLIIPWLDYLKKRRNRNICLKEF
PMIPCTSOVMTQAITFEYASSISIFAFCVLVAFITDSFQNVFLHSSYLKPLIKI
SGSMSSHESGMLLMFCILQIVSCYIECEDKMKFSSIIILSAIOLFFSFYETSN
PFHMFSPVKEGILNPMLODIALSIHPPLLYIGVYAVPTIIVCAIFLIAVISRLD
NMLPGYCTGKTKQLSDRTSIRLIKIFACGILFTTIGISLSMWAIRLNGGLME
FDVENISLEFPMISGIMLHSITVTRKNMOWTIIISITFLVIFSTFLVRSGET
NSIHAFESSEKWIYLVILCIIGIGLVYIIRLNCSIEKTIYNFTNMWTHSCNWT
IKKAILIGNIFELSLILICATLVPTIYALFNDKPIINEEFIKNFIFVPIIC
TILLETNNSIKKIHIMLMSLITIVFIILNIKFSVSIILTIMTSIFIMHNYILV
KTYFERNLSKASTSMILGHFGVALLAFSTNMSLQSEMDIGEVGTSFPNAPFYT
LQNIKFAQGNKYROIJAEEVLEDKNNKITLKRPENRYITTEOKLSQESDHSYLVLDL
YAVLSNDGKIHKATIKYKPMASFIMLVGVLITTAGVIALIRKNSY"
gene
complement(7264. .12195)
/gene="Rp704"
complement(7264. .12195)
/gene="Rp704"
/codon_start=1
/transl_table=11
/product="CELL SURFACE ANTIGEN (scas)"
/protein_id="CA15140.1"
/db_xref="GI:3861241"
/translation="MAOKPMFLKKIISAGLVASTATYVAGSGVAMGAMONRTTN
AAATPDDGCGDOAGANIPVAPNSVITANANPENTPENGHLSIFLTDALAVT
INBDITLGIITTAQOKAFNFYVAAGKILINITGOGITVDEASNTITNAQALIKVGG
AAINANDLSGISITFAAPASVLEFNLINFTQDEAPLTGLANSKIVNGGGLNTING
FIVQSDNTEFAGIKTINIDDOGIMFNSTPDANLTNLQVGNNTINENGIDGTGLVLY

SKNGATEFNVTGLGNLKIIEILNTAAVAGKLISOGAANAIVIGTNGCAGRAGPI
VSYDNGMAATISGOVYAKNMVIGVSAANGGVFEHHIVDVGLGTETPKRDSVITTE
NSNPGSTNRENLDQIYVPEPTKILKGFIDGVAKNNGTAVITFPMNGALYSSTDPN
IATVNTALIEABGAVVELSGIHAELRIGEDGAYKNNGTAVITFPMNGALYSSTDPN
LAAGSIQLODGSAAITDIDGNGVNAALOHITLVAADSKIALLDGANILGAVGALHE
QANGGTIKLNTQNNIYVNEFDITDITDGTGVNDASSLITNNQTLINGSISTVANTKT
LAQNLGSSKTIINAGDVAINEIVENGSVOLNHNRYLTITKTITANAGQILIYAADP
LANTITLADGTNINAGSAPNPLSTIHPATKANAGSILINVGKVALYNNITNTNANGS
LHFRSGSTISVSTVGQOCHKLNNLILNDGTYKPLGDTENGKRIEBSLIQJLSN
NVTTHVESADNQTGLEFVPTDPIYVTLNKGAVFGYKQVITISGPNIVFETIGVG
IVHGLIANSISIFENASLGTSLFLPSGTPDLVLIKSYNGVGVNDNAPLVAVGSLDS
MINNGOIIIGDKKNIILSLGSDNSIYVNAATLSGRTTNNQGTVALISGMPNNGT
ITGILGELGSPKIKOVYFTTDYNNLGSILNANNYITNDYVLTGAGIAGTDFPKITLG
SVNGANVRPVDSTFSPRSMIVATQANKCTVYIGALVANSITGSDTTPYASRFTGN
DSGAGILOGNTYSNIDEGYNTLITNSNVLGGGTAININEIDILLNLIIFANGSTW
GNDTISTITLNVSSGNIGOVYIADQVANAITGGTITTKIADQANANFSTQAVTLIO
GGARFNGTGLAPFAPEAVTGSNIPEYELIRSDNDVYFALITDPSAAVTTVAAGSALAN
APGVSONISCELESTMTAAVNNMLLAKDPDVAFTFGALITDPSAAVTTVAAGSALAN
QDLSNRLGTRILSNAETSDVAGSAGVAGSDEAEVSVGWAKEFYVINAEDOKKG
IAGYKRAKTTGVVYGLDILASDNLMIQAIGITTKTDIKHODYKKGDKTDINGLSFLYG
SQOLYKNTFPAQNAITFLNKYKRSQRYFPESGKRSKOIAAGNYNMPGSLITGCG
DIVAMPVILVTPMAGLSYLKSSNENTKEGTITANKRINRSDDRDILVGAVVAGST
VNTDILVIYPEIHSFVHKVNGKLSNSQSLDQGTAPFISQPDRTAKTSYINGLSANI
KSDAKMEYIGIGYDFNSASKYTAHQGTLKVRVNE"
complement(12953. .13216)
/gene="Rp705"
/note="Pseudogene that lacks ATG or GTG initiation codon.
Homology to guanosine-3,5-bis(diphosphate)
3-phosphophydrolase."
complement(12953. .13216)
/gene="Rp705"
complement(13563. .14507)
/gene="Rp706"
complement(13563. .14507)
/gene="Rp706"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA15141.1"
/db_xref="GI:3861242"
/translation="MKIRQYAKPVYIIGSPBELTDAERLEFENHPLGIILFRNRIRK
NEKGEQDKAALIKLADIKEVLGDNITIIAIDOGGAWKRIAPFYDAPAPQFTBLQ
TCKYNSYTIKELREVGWINDPFVADLHDGDAKLSIDRSFKEBEIYVFLPSAIA
GLQEEKYATCIKHIHPGSRATVDSHLEPIIDGSLTLEPTDFKVRKELAKYDIKIA
MTAVYIYALDPPNPATLSQIYDIYKSNIGFGLIISDITENQALSGSMADITKRAL
DAGVDIVLECTGEFNMKSEVYLSVAERSIDKFSDFLTS"
complement(14552. .15025)
/gene="Rp707"
complement(14552. .15025)
/gene="Rp707"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA15142.1"
/db_xref="GI:3861243"
/translation="MINNKLTKRYVASEVYRILNIALHKLRYLETKIGLSHYKINN
KRYTARHDLDLQFLVRSREYLYRGVHRSDXQYLAHNSIHGASDPSRNEOG
KNSCELTINLHVQORODSQSAVYVSLMDNVTSSRIDLTTFNHLISIQIKKILLIS"
complement(15220. .15507)
/gene="Rp708"
complement(15220. .15507)
/gene="Rp708"
/codon_start=1
/transl_table=11
/product="INTEGRATION HOST FACTOR ALPHA-SUBUNIT (hima)"
/protein_id="CA15143.1"
/db_xref="GI:3861244"
/translation="MTITKNIKISLMSKLGFSNNLCEAIVNTVFSNILEIAVOKLT
LKNFSFEYKQKPPRGINHTKSPVWASKKRLRTPSPSKKALINKSML"
18405. .18701
/gene="Rp709"
18405. .18701
/gene="Rp709"
complement(18701. .18701)
/gene="Rp709"

```

/codon_start=1
/transl_table=11
/product="unknown"
/db_xref="GI:3861245"
/transl_table="NAVKQYIEDHTEDLHKICNLVKYIELETAITHEKAEDTEV
VIEVSLIDNSSDDASIIINKKAITVTSDDIANSNEDAAINTKSHQDYLRCYQ"
complement(19685..20789)
/gene="RP710"
complement(19685..20789)
/misc_feature
/gene="RP710"
complement(19685..20789)
/gene="RP710"
/note="Pseudogene with two authentic frameshifts and four
authentic in-frame termination codons. Homology to
transposase."
complement(21112..21608)
/gene="RP711"
complement(21112..21608)
/gene="RP711"
/note="Pseudogene with one authentic frameshift. Homology
to resolvase."
complement(22602..23065)
/gene="RP712"
complement(22602..23065)
/gene="RP712"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA15145.1"
/db_xref="GI:3861246"
/transl_table="MSIEKEKFMASHKQYVEIGETNLINGKHLDFKAVATSDIEA

Query Match      9.3%  Score 45;  DB 1;  Length 237523;
Best Local Similarity 46.9%;  Pred. No. 1.4;
Matches 175;  Conservative 0;  Mismatches 195;  Indels 3;  Gaps 1;

Qy  100  GGTGGCGTTGTGGCGCGTGTGGCGCCAGCTGTTGGTAAGCTCTGTCGTGTCG 159
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 172645  GGTGGCGCAGAGAGTGCATTACTTGTTGCATTCGTCAGTAAGCAACAATTGTC 172704
Qy  160  ATGGCCATGGCGCGTCCGCTTCTGGCGGCTCTATTGGCTTAAATCGTCAGACATG 219
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 172705  GGAGTGGTGTAGCGCATTAATCTTGGGCGAGTCTTGCGGACAAATCGCGCAAGTATG 172764
Qy  220  GACCCAGGATTAATCAACTGACAGCAAGTCTCTGGAAGAAAGTGAAGCGCCAGGTT 279
      ||  |||  |||  |||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 172765  GATGACAGAGATGAGAACTTCTAGAACTAACATTCACAAAGACCTTTA---GAATCTGCA 172821
Qy  280  ACTCGTTGGCGTAATCCGACACCGGTAACTGCTGTGGACCGGTTCCACCTAC 339
      ||  ||  |||  |||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 172822  CCTAGCGGTAGTAACTATGAAATGGCGCAATTCAGATACGGCAATCATGTTACGTCACA 172881
Qy  340  CAGCGTTACAAACAGACGACCGCTCAGCACTACTGCCGCGAATTTACGAGAAAGCC 399
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 172882  CCTAAATAAACTATATGGAACATGCAAGTCAATATTGCCGTGAATACACTCAACAGATT 172941
Qy  400  ATGATGCGAGGTGAGAAACAGAAATCTAGCGGACCGCGTCCAGCCGATGGCCGC 459
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 172942  ATTAATAGCGGAGAAACAAACAAATATGATGATGATGATGATGATGATGATGATGATG 173001
Qy  460  TGGCAGGTGATTA 472
      ||||  ||  ||  ||
Db 173002  TGGCAAGTTGTGA 173014

RESULT 5
AL646071  203050 bp  DNA  linear  RCT 07-DEC-2001
LOCUS     Ralstonia solanacearum GM11000 chromosome, complete sequence:
DEFINITION
segment 15/19.
ACCESSION AL646071 AL646052
VERSION    AL646071.1  GI:17429627
KEYWORDS   Ralstonia solanacearum.
SOURCE

```

```

ORGANISM
Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia
REFERENCE
1 (bases 1 to 203050)
Salanoudat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brotier,P., Camus,J.C., Catolico,L.,
Chandler,M., Choisme,N., Claudel-Renard,C., Cunha,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,
Siglier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Unpublished
2 (bases 1 to 203050)
Boucher,C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian BouchetToulouse.inra.fr/R.solanacearum.html.
http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES
source
CDS
gene
complement(180..1025)
/note="RSC2606"
/gene="RSC2606"
complement(180..1025)
/function="elements of external origin; plasmid-related
functions non-bacterial functions"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon-usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="HYPOTHETICAL REPLICATION INITIATOR AND
TRANSCRIPTION REPRESSOR PROTEIN"
/protein_id="CAD16313.1"
/db_xref="GI:17429628"
/transl_table="MSSPPGQALPQREQLDFRALPGDMPAPDSODLMAFPFSLAKS
RRTAPIDFRAGGITRVEGTAGHGIATWDADVLIMASQIYEARDAGIRPSRWIRAT
PYEILRFIGRSTLNDYORLKAJLRLOSTSTSIRETGRLHRSFVINGRRLAD
ASGKPLGELIILPDWEYAGVDAALVITDIPYFLKGIIEMLRLIVKKGROEHG
WQDFRKLHHRKGSARSDFXIDLALVARSLSGYVLGIERMDGGCELLTFRPVL
HTAGC"
complement(1052..1327)
/gene="RSC2607"
/note="RSC2607"
complement(1052..1327)
/note="RSC2607"
/gene="RSC2607"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon-usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD16314.1"
/db_xref="GI:17429629"

```

gene .
/translation="MRPAPLRPAAPAPQPORRYLTINPEAAYLRLSPRTLEKQKRVGS
GKRFKFKGRRWYAVASDDANADQNSYATSDPEYAEHNGADYDGR"
complement(1442. .2212)
/gene="RSC2608"
/note="R800857"
complement(1442. .2212)
/gene="RSC2608"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HYPOTHETICAL/UNKNOWN PROTEIN"
/protein_id="CAD16315.1"
/db_xref="GI:17429630"
/translation="MTDSAEHWYPTAYLYVLHIDGEPALAMETLRRNPQYRRMLRR
RRRPDAQWGLRLLEDPALDRADNRAMPFPDHDVVOLFPDADPEEHADEFWRVP
GKKOLIHDKRLVIVSHWPGCLRLALAPGLEDMAYLYATACATPCARLTJATL
DALATAYVAPTPAARSPTLAAVLEHTTQALDATALAGASIREVAAGLPSADAVAAD
HKWDSLRLRRVRLVRGEGALMRGGYRLQLPILPIQ"
2213. .8108
misc_feature
/note="ACUR or RS00937
predicted by Framed"
/evidence=not_experimental
complement(2557. .3150)
/gene="RSC2609"
/note="R800858"
complement(2557. .3150)
/gene="RSC2609"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE LIPOPROTEIN"
/protein_id="CAD16316.1"
/db_xref="GI:17429631"
/translation="MTERPLSCFYGNPHLRIPTEGGRIRIRFASDMSASVLPFGCC
DLIHTRASTRMIRGIVLCFTGRAPRIVGVFRLPPTLPVLADERARLHAGAL
AGGRIDPLVVRLETPDAHVTLVLSLDPDGDTAGLIDLGIGLPSLGTGVKXLSDLA
GIVGPRKLEPVRDRYFOAHHLLSEYVRLAONGAITD"
3223. .3513
gene
/gene="RSC2610"
/note="R800859"
3223. .3513
CDS
/gene="RSC2610"
/function="miscellaneous; not classified regulator"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD16317.1"
/db_xref="GI:17429632"
/translation="MAAKYSLAKALKTVRKARGLSQEAFSVSSSRTYASTLERDLKSP
TLHKLTKEISQVNDVHPLTLTLTAFAGDSTRADELLAQVRELAIVLNSDD"
3522. .4085
gene
/gene="RSC2611"
/note="R800860"
3522. .4085
CDS
/gene="RSC2611"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage"

Query Match	Best Local Similarity	8.9%	Score 42.8	DB 1	Length 203050
Matches 80	Conservative 1	0	Mismatches 62	Indels 0	Gaps 0
QY 327	GGTGGACACCTTACGACGCTTACCAACAACAGGACGACGCGTACGACGATCTGCGCGGAAAT	386			
DB 105900	GCTTGCCGCTCTGCGGCTTACGACGATGCGGACGACATCCCGTGGCAGCTGCGGCGCAT	105959			
QY 387	TCAGCAGAAAGCCATGATCGCAGGTTCAGAAACAGGAAATCTACGCGACCGGCTCCCTCA	446			
DB 105960	GGTGGGCGACGCGGATCGACGCGGCTGATGACGACGATGACGCGGACGCGGCTGCTCA	106019			
QY 447	GCCGATGCGCGCTGCGACGATG	468			
DB 106020	GCCGACGCGACCTGCGCAGATG	106041			
LOCUS	RSU76907	448 bp	DNA	linear	BCT 18-FEB-1998
DEFINITION	Rickettsia sp. 17kDa common-antigen gene, partial cds.				
ACCESSION	U76907.1	GI:2894794			
KEYWORDS	Rickettsia sp.				

ORGANISM Rickettsia sp.
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsiae; Rickettsia.

REFERENCE 1 (bases 1 to 448)
AUTHORS Davis, M.J., Ying, E., Brunner, B.R., Pantoja, A. and Ferwerda, F.H.
TITLE Rickettsial relative associated with papaya bunchy top disease
JOURNAL Curr. Microbiol. 36 (2), 80-84 (1998)
MEDLINE 98087556
REFERENCE 2 (bases 1 to 448)
AUTHORS Ying, Z. and Davis, M.J.
TITLE Direct Submision
JOURNAL Submitted (01-NOV-1996) Tropical Research and Education Center,
University of Florida, 18905 SW 280 Street, Homestead, FL 33031,
USA

FEATURES
source Location/Qualifiers
1..448
/organism="Rickettsia sp."
/db_xref="taxon:789"
/note="Puerto Rico"
CDS
/note="putative"
/codon_start=3
/transl_table=11
/product="17 kDa common-antigen"
/protein_id="AAC02809.1"
/db_xref="GI:2894795"
/translation="AATAASMLEACGPGGKMGKSGTLIGTIGLVSGFGGGRRL
AAVAGALGALIGNIGAGMBDQKLELTSORLEAPSSSSVOMNPDMNGT
VTPSKAYKNNTQYCREYQTVVYVGGKQKAVGTACRPDGGQVY"

BASE COUNT 135 a 81 c 120 g 112 t

ORIGIN

Query Match 8.6%; Score 41.6; DB 1; Length 448;
Best Local Similarity 47.7%; Pred. No. 8.3; Mismatches 199; Indels 3; Gaps 2;
Matches 184; Conservative 0;

87 TGGCGGCGCCAGCGGTGCGGTGTGGCGGCGGTGTCGCGCCAGCTGTCGTAAGGCTC 146
Db TGGTACCGCTAATTGGCGCACTTTAGAGAGACTGTTGTTGCACAAATTTGGTGTGTAC 124
QY 147 TGGTCTGTGTGATGCCATCGCGGTCGGGTCTGGGCGGTCTGATTTGCTTAAAT 206
Db 125 GGGACGACTTGTGCGGTAGAGACGAGTGCCTTACTGTGTCAATCCTTGGCAACCAAT 184
QY 207 CGGTCAAGAGATGAGACAGAGATAAATCAAACTGAACAGTCTCTGAAAAAGTGA 266
Db 185 TGGTGCAGTATGATGAACAAGATAGAAAACTTGACAGACTTACTCTCAAGAGCTTT 244
QY 267 AGCCGCGCAGTTACTGCTTGGCGTAACTCGGACACCGGTAACAGTACTCTGTGAAC 326
Db 245 AGA-ACGTGCAGCAAGAGGTGTACAGTGTACAAATGCGCTAATCCTGATTAATGTAATG 303
QY 327 GGTTCGACCTTACAGCGTTACAACAACAGAAACGCGTCAAGCAGTATGCGCGCAATT 386
Db 304 GTACTGTGACACCAAGCAAAAGTTATTAATAATATACCG--GTCAATATTTGCGTGATA 361
QY 387 TCAGCAGAAAGCCATGATTCGAGGTGAGAAAGAAATCTAGAGCCAGCCGTCCTCA 446
Db 362 CACCAACAAAGTTGTAGTAGTGAATAAAGAAAGTTATGACACTGCTGCGCTCA 421
QY 447 GCCGATGCGCGCTGCGAGTGATTA 472
Db 422 ACCTGATGACATGCGCAAGTTGTGA 447

RESULT 7
AL646061
LOCUS Ralstonia solanacearum GM11000 chromosome, complete sequence;
DEFINITION segment 5/19.
ACCESSION AL646061 AL646052
VERSION AL646061.1 GI:17427781
KEYWORDS

SOURCE Ralstonia solanacearum.
ORGANISM Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.

REFERENCE 1 (bases 1 to 198050)
AUTHORS Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S.,
Arlat, M., Billault, A., Brotlier, P., Camus, J.C., Catolico, L.,
Chandler, M., Choise, N., Claudel-Renard, C., Cunac, S., Demange, N.,
Gaspin, C., Lave, M., Moisan, A., Robert, C., Saurin, W., Schlex, T.,
Siguer, P., Thebaud, P., Whalen, M., Winkler, P., Levy, M.,
Weissenbach, J. and Boucher, C.A.
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 198050)
AUTHORS Boucher, C.A.
TITLE Direct Submision
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, UMG CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Bouchere@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.

FEATURES
source Location/Qualifiers
1..198050
/organism="Ralstonia solanacearum"
/strain="GM11000"
/db_xref="taxon:305"
151..642
/gene="RSC0770"
/note="RS05084"
151..642
/gene="RSC0770"
/FC_number="1.-.-.-"
/function="small molecule metabolism"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="PUTATIVE TRANSMEMBRANE DEHYDROGENASE (SMALL
SUBUNIT) OXIDOREDUCTASE PROTEIN"
/protein_id="CAD14472.1"
/db_xref="GI:17427782"
/translation="MHTPPSSDGDPAQITRROWLSALAAATPAASLVGSVTLVAYAQ
AGCEPLDAFMGSQLATARSALDRAAGALALADASGFAAQLPLAQSILAAGTLS
GQGEETALRIIQAWYGVVDGVTVEALMGVYSDIIVITRYCPNRGFAEPVE
KOS"
642..2261
/gene="RSC0771"
/note="RS05083"
642..2261
/gene="RSC0771"
/FC_number="1.-.-.-"
/function="small molecule metabolism"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="PUTATIVE TRANSMEMBRANE DEHYDROGENASE (LARGE
SUBUNIT) OXIDOREDUCTASE PROTEIN"
/protein_id="CAD14473.1"

gene
CDS

gene
/db_xref="GI:17427783"
/translation="MAETRQAEQADIVVSGVAGALVAYELARAGKSVLMLAEAPRL
PRMEIVERRQADKMDPMAPYPTSPNAPHEVGPDDNNYLILKEHEFDSQYIIRAVGG
TTHWAAATPFLPDEKFLRSVYGIARMPLOYDILREYGRARALAGVGNDEELG
SPRSQYMAPMLPLSFNERETKVDALNMHPDLRVTEPVEVANSRSPYGRPLCGGNNG
MPICIGMNYNGIYHEKARQAGARLILEMAYVYELVAGACRIYAAIKYKPKQKDHVY
EGKMYLWANGILETKLMLSTGADPFGVGNESDMDGRNIMDHGTGVSFYADRKIM
PGRGQENTSLIGFEDDGFRAVQAGKLIHLNISIRICQETTRIFERAKKLIKPAIDAR
IRDOARAVQSDSFHEILPLPENRIVPSATGTDALGIPREITRIDDYKRSVHVR
EYVAARAVQSDGVQFDHDPAPNNHITGATIMSDPKDSVDCDKRTFDPHNLFISS
SATMPTVTVNVTLITALLARIADQLKKEA"
2263. 3552
/gene="RSc0772"
/note="RS05082"
2263. 3552
/gene="RSc0772"
/EC_number="1.-.-.-"
/function="small molecule metabolism; energy transfer;
electron transport"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="PUTATIVE OXIDOREDUCTASE DEHYDROGENASE
(CYTOCHROME C SUBUNIT) SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD14474.1"
/db_xref="GI:17427784"
/translation="MRRKASFFLLSVLGAANAARDAPADAQAIRKGEYLAIAGD
CYACTATAGGKRFAGKPLQIPIGLGTIVSSNITPDPYIGKMYEDFEERAVRGVSD
DGENIYPAVGSYAKIINDADMRDLIYEFRGVAOVAPPGPIITMPLNNMRKTLKLA
NIARLKRAPIYLKGMQPTNRGATLACGLAHCSTGCRPGFAOEKFAOERGGYLA
GSTLGAESYNTISDPVSGISGMKEQDVOYLTRVPSGPGICQANGPGEAVOHFSFSM
TERDVARLAEYLRITPAVASTGAORARHSGWPAIDVIALRKPLETEIDARLYLVGSK
ATGHQDAPGPDGYPRPLNLSAVGADFTNLVQVMLHGIERRAGDRHITGMPAFGRQ
LSDAQLAALANVTYRFQDPPATPALTADEIAKRRMPQ"
complement(3559. 5781)
/gene="RSc0773"
/note="RS05081"
complement(3559. 5781)
/gene="RSc0773"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="PROBABLE TRANSMEMBRANE PROTEIN"
/protein_id="CAD14475.1"
/db_xref="GI:17427785"
/translation="MRTSARVILGLTPIPIVAAALVFLTDMNBRAPRYLNDHVSQ
ANGRFALINGDALTWKKRDESGSGWRVAPVPRILASDILGNDMDKROPALYNDHVSQ
TRYEALPLLAHRIYIPSVTIDTPKILALERSQNRNNMTFHGQDSDSPSKLELEK
IAFSGQVQLDDEVKRIHLAATLVDVQALVYASGTAIGSSPEPAASAGAASTRA
PAAPDPOPYGIAMTVKCTYNNALIRGGKAGVYLRLQDARPPRLADAVYNGATRID
LAGLTITNSSTLADLRILHLSGASMAHLYPLTGVLPTPEPDRGRILIGLRGSSSS
WMEYDFGTGSSDGLTFLTAMREPEQRORLGEIASHOLFLPADLAPITGAOSNNS
KORRAPPARQADKYLIVPEPRTDWMANIDAVYATRYRTADLPIDHLYIKTIO
DAVLTLDPLNFSVAGCTILSNLRDQSGAPAPQASALAAHRIKROLFPNTESHAIST
GAINDARLASIGNEVALILGASNGEKLIVEQGVSEFILLEAGLNGVNTLTKLFGG
DKVYSINIAAGGFVASDIAQARTFVVDTOAVIDYGATSEKESIDFTIHPDSKD
RFVYSRTPLVYGYTKHBDVSNPAVVALRAGGAVALAFAPVAAYLVLETLQAPDSQ
PCGLKLADVRRQPTAPPPGKYRSGQGRADGSPKDPAAAPARQPALPLMDPTTAGG"
5683. 6785
/gene="RSc0774"
/note="RS05080"
/db_xref="GI:17427786"

Query Match	Best Local Similarity	8.5%; Score 41; DB 1; Length 198050;
Matches 83; Conservative	0; Mismatches 70; Indels 0; Gaps 0;	
61	61	61
121	121	121
135979	135979	135979
181	181	181
136039	136039	136039
AL646079	AL646079	AL646079
LOCUS	LOCUS	LOCUS
DEFINITION	DEFINITION	DEFINITION
ACCESSION	ACCESSION	ACCESSION
VERSION	VERSION	VERSION

KEYWORDS	Ralstonia solanacearum.
SOURCE	Ralstonia solanacearum
ORGANISM	Bacteria; Proteobacteria; beta subdivision; Ralstonia group; Ralstonia.
REFERENCE	1 (bases 1 to 213050) Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S., Ariat,M., Billault,A., Brotier,P., Camus,J.C., Cattolico,L., Chandler,M., Choise,N., Claudel-Renard,C., Cunnac,S., Demange,N., Gaspin,C., Lavey,M., Moisan,A., Robert,C., Saurin,W., Schlex,T., Siglier,P., Thebaud,L.P., Whalen,M., Wincker,P., Levy,M., Weissenbach,J. and Boucher,C.A. Genome sequence of the plant pathogen Ralstonia solanacearum
TITLE	Unpublished
JOURNAL	2 (bases 1 to 213050) Boucher,C.A. Direct Submission
REFERENCE	Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
AUTHORS	Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
JOURNAL	Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, DMG CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT	Christian.Bouchere@toulouse.inra.fr http://sequence.toulouse.inra.fr/R.solanacearum.html. Location/Qualifiers
FEATURES	1. .213050 /organism="Ralstonia solanacearum" /plasmid="megaplasmid" /strain="GM1100" /db_xref="taxon:305" 100. .1632 /gene="RSP0487" /note="RS00365" 100. .1632 /gene="RSP0487" /function="macromolecule metabolism; macromolecule degradation; degradation of proteins, peptides, glycopeptides" /note="Product confidence : hypothetical Gene name confidence : hypothetical Predicted by Codon_usage Predicted by Homology Predicted by Framed" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="HYPOTHETICAL TRANSMEMBRANE PROTEIN" /protein_id="CAD17638.1" /db_xref="GI:17430957" /translation="MDAPPSGAVTTRAVGRLLVLYVGLAVLAVCIAPFNTRRHDSR OUDVPAAPLAVDGAARAAARIGEARIRATISSPTDAONADOFROHMLAARYPAH AMQREPVDSLSLTWKSDPSLRKILLAMADVYPVAPGTGDTPEPPAGVVDG KWWRGAMDKGLIAQMEAEILASGFRPRRTTHFAFGADEVGEGERAAITALL KSREGDLAVIDEGGLITEGVVGLRKPAPALIGYBEKFLSVALKLSAPGHSMPA PESAIAMMSAALKHLDDQOLPAGIRGVREMEETLAPEMRGINRYALSNLMFEPLV RQOLASPTSNVLTOTTTALTIVQAGNKDNVLPGRAEATVNFRLPGDSASSVIAVE OAVRSAPKGFHEIALPGVSEAPVSPQASAYOLIGRTVREVPGRVAVARIMGA TDSRHMITGISDHFREFSPVAPRDELARFHGHNRISEANVLELIRFYRLHIOADA GA"
gene	complement(1751. .3148) /gene="czcs" /note="RSP0488; RS00366" complement(1751. .3148) /gene="czcs" /BC_number="2.7.3.-" /function="cell processes; protection responses" /note="Product confidence : probable Gene name confidence : probable
CDS	
gene	predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="PROBABLE TRANSMEMBRANE SENSORY TRANSDUCTION HISTIDINE KINASE FOR COBALT ZINC CADMIUM RESISTANCE TRANSCRIPTION REGULATOR PROTEIN" /protein_id="CAD17639.1" /db_xref="GI:17430958" /translation="MMMPASLSAARLVAFVAVAGVMTGAGTYLYVLAQLRATD AMLKIVOLRHCAKDQSIDDIVDARRETVVSGHEGLIRVIGPDRTVEANPE GEVAPAPAPVPDVLASRESALHTWMSRCSIAARATVAAQRLGNHRTDVTVAQOGDS RAIMASVTRHIFLAVIACVAAVALISIMVSTIKGARTGROASVYSPDTRLT LEGAPRESDLLAALNMLTRLDGGRARLSQPSADLAHDFRTPLANLIGQETVLAP RPAQERAVLESSELEEARLSRIDMLEFLARADARVALSVAPLARHDIQVLDF EVAADRDITIOVCGGRVADLTLLRAINLNLNLRHTPGARIDVAVQVLPST ATAVRNPGGTPAASLPLEFDRFYRNDPARANSERSAGLAIYQTIIMDHGTAQAS SPPGUTFERLTFPA"
CDS	complement(3145. .3819) /gene="czcr" /note="RSP0489; RS00367" complement(3145. .3819) /gene="czcr" /function="cell processes; protection responses" /note="Product confidence : probable Gene name confidence : probable Predicted by Codon_usage Predicted by Homology Predicted by Framed" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="PROBABLE RESPONSE REGULATOR FOR COBALT ZINC CADMIUM RESISTANCE TRANSCRIPTION REGULATOR PROTEIN" /protein_id="CAD17640.1" /db_xref="GI:17430959" /translation="MRLLIVEDEPKRAGDYLKGLRESGVADLARSQPGLYOMEND YDILVDVLMFGMDGWOVIRELRRRTTVVLELWADSLDRKLGLEGADYLVKPF APAEELVARHTILRGRPMRESEVLEIADITITIRKRVTRARRIDLTAKEXVALHL ARRTEGVLSRSLISSQVMDVNFDSNTNVDAVIRRLRAKAVDDPFEHKLHTRGNGYV LDTVAR"
gene	4089. .4394 /gene="RSP0490" /note="RS00368" 4089. .4394 /gene="RSP0490" /function="miscellaneous; hypothetical/global homology" /note="Product confidence : putative Gene name confidence : hypothetical Predicted by Codon_usage Predicted by Framed" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="PUTATIVE SIGNAL PEPTIDE PROTEIN" /protein_id="CAD17641.1" /db_xref="GI:17430960" /translation="MSVRNCLLPFLAAAFACGAATAASYEQALSPLOYHGARDP YNDGARGCTRDPTPDGAQAEATLQASVGRPDYSDGMSRSSAPRRDRDPTDGAQAA" 4400. .5725 /gene="czcg" /note="RSP0491; RS00369" 4400. .5725 /gene="czcg" /function="cell processes; protection responses" /note="Product confidence : probable Gene name confidence : probable Predicted by Codon_usage Predicted by Framed" /codon_start=1
CDS	
gene	predicted by Codon_usage predicted by Homology predicted by Framed"
CDS	

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	SOURCE	
RESULT 11	AEO08675	10127 bp DNA linear BCT 14-SEP-2001						
	AEO08675	Rickettsia conorii Malish 7, section 107 of 114 of the complete genome.	AEO08675	AEO06914				
	AEO08675.1	GI:15620413						
		Rickettsia conorii. Rickettsia conorii Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group. 1 (sites) Ogata,H., Audic,S., Barbe,V., Artiguenave,F., Fournier,P.E., Roult,D. and Claverie,J.M. Selfish DNA in protein-coding genes of Rickettsia science. 290 (5490), 347-350 (2000)						
	MEDLINE	20485642						
	PUBMED	11030555						
	REFERENCE	2 (sites) Ogata,H., Audic,S. and Claverie,J.-M. Selfish DNA and the origin of genes science 291 (5502), 252-253 (2001)						
	TITLE	Science 291 (5502), 252-253 (2001)						
	JOURNAL	Science 291 (5502), 252-253 (2001)						
	REFERENCE	3 (bases 1 to 10127) Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V., Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. and Roult,D. Mechanisms of Evolution in Rickettsia conorii and R. prowazekii Science. 293 (5537), 2093-2098 (2001)						
	TITLE	Science. 293 (5537), 2093-2098 (2001)						
	JOURNAL	Science. 293 (5537), 2093-2098 (2001)						
	PUBMED	11557893						
	REFERENCE	4 (bases 1 to 10127) Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V., Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. and Roult,D. Direct Submission Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine, CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean Moulin, Marseille Cedex 05 13385, France A public version of R. conorii genome database is accessible at http://igs-server.cnr-mrs.fr/ . The database intends to provide updated data. Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful. Comments to the authors are appreciated. Location/Qualifiers 1..10127 /organism="Rickettsia conorii" /strain="Malish 7" /db_xref="taxon:781" 111..890 /gene="map" /note="RC1276" 111..890 /gene="map" /note="REP02, RS3-like repeat" /codon_start=1 /transl_table=1 /product="methionine aminopeptidase [EC:3.4.11.18]" /protein_id="AA03814.1" /db_xref="GI:15620413" /translation="MTIKIRHGKGFIMRAAGKLAENFDITPDHVKPNVTNSLNGI CHNRTSHNAVPAFLNKPKFKSCTSTINHVCIGIRDRKLKNGDIYNDIVYIIIDGQ WYGTSRMRYVGDVAIKPKRLIQVTDMMKGILEVVRGARLDIGHAIOSYAEKHMYT SVVDYDTHSGISAGFEHTPTILNGRGTGTLLEGMFETVEPVLINGANDTILSKLDD GMVTPRDKSLSAQFERTIGVTCKGEFIPLSPKKLYPPY"						
	CDS	111..890 /gene="map" /note="REP02, RS3-like repeat" /codon_start=1 /transl_table=1						
	gene	111..890 /gene="map" /note="RC1276"						
	CDS	111..890 /gene="map" /note="REP02, RS3-like repeat" /codon_start=1 /transl_table=1						
	FEATURES	source						
	COMMENT							
	gene							
	CDS							
	repeat_region							
	repeat_region							
	gene							
	CDS							

gene	CDS	GGGTDAVLAATAGRRRCRYEGPLGEPTLAHGMWLDATAVLEMASAGSLHVPDR RDATRRSSGRTGELIRALDNGARKIILGLGSGATSDAGAGLGLVPLADDEBL APGGAALAGLSLDLGGDLPRLVDVAENVADNDPLCGPRGASVGPCKASAEV AOLDAALAHFAKVAATLDEDFSVPGVAGAGLGAARAPFARPRRTEILVALAG LADALVAGDVLVTEGRDLDAOSLHGKTPVAVIARQACVPVVALAGSLGDDYORLHD AGIDAFSLVAGPISLEOCACGAAEIENARQIARLMLAAGASREG* 7313..7777 /gene="PA1053"
		7313..7777 /gene="PA1053" /codon_start=1 /transl_table=1 /product="conserved hypothetical protein" /protein_id="AAG04442.1" /db_xref="GI:9946967"
gene	CDS	TTGALRPKIEGTPTPIGSIAGAGVAGSAGSAGSVAAIIGAVAGLIGATE EGILTRGQVEIVREDSDSTRAYVQVDDGQIFVGERVHILTVNGTSRYAH* 8106..10907 /gene="PA1054"
		8106..10907 /gene="PA1054" /codon_start=1 /transl_table=1 /product="probable NADH dehydrogenase" /protein_id="AAG04443.1" /db_xref="GI:9946968"
gene	CDS	GORGVFAGQITRTASTSWPEIGINLSLRDGIQFALLIIGLITVITARTLSE OEMRGEPARILLPMSMIGVLAENLMLFWEITSLSSPRLIGLITVITARTLSE PMATVYGGGLLAGVLLGHIGSFELTVEVLAAGELIRAHALPLVLLILGVF TKSAQPFPHMLPMAAPPVASVLSHTAKVAPILARIPALVLLILGVF SGVTLILGAGMALFOHDLKGLAVSTISGLITLIGLSDPLSVAVAHIINHT FKASLEMAAGIIDHETSRMDRLINMMKXYDHTVLAIVASMAQVPLNGFSE MFEQELSONLIGSPMVLPAVATLAGVSASISRETHDFEFGEEARLPKPEP PRYAKPVEVILVCLVGMPLPAYAPILAAASASIGQIPXSIAIMGFVPLL MSYVALGVLYLVRKPLPNYACLPEDKLVLEDQOVVALAARLTAENGST ORYLAMLGLAVVAVELAPARLTSRGTLPLDGLTALGLVLAFLASGLATPIRM RIVALLISVAGLVAVELAPARLTSRGTLPLDGLTALGLVLAFLASGLATPIRM SURGLRDVLAGVGLVALAVLTPREYSIAGFEVENSAGGGSNNVNVILDF RGPDRTGETISVLAGVIVAMLAGFALHFKPTCDPOGRNNAHAKHPLILETSRVLDF LALLISVFLPLRGNHLPGGGFAGLVTAIVLYOITVAGSOWEORPLNVAMAGAG VILAGTILSGMSLFCYPLTSAFGHEPLIGELIETALPLDGLVLYVAGALLIL STLGLSLMPEPVAAEAL* 10907..11245 /gene="PA1055"
		10907..11245 /gene="PA1055" /codon_start=1 /transl_table=1 /product="conserved hypothetical protein" /protein_id="AAG04444.1" /db_xref="GI:9946969"
gene	CDS	AMGRGSSVAVIGTPGEYGDPLDQALVLAIVIGFAMAFVVLRLGLDGLSDHD GRPEEERPR* 11242..12741 /gene="PA1056"
		11242..12741 /gene="PA1056" /codon_start=1 /transl_table=1 /product="probable NADH dehydrogenase" /protein_id="AAG04445.1" /db_xref="GI:9946970"
gene	CDS	LIIRTAGSDLSVALGNMAMPFGLVLMRLDRLAIMLTAIVLSGSAIILYALRGEDRL GHRHALFOQLINGALITGLDFNLFPFPELLITASYALILHGGAGRVSGGLHY IINTLVGSAFFLIAGTIVGLTGNMADMAOKIAMDADERAPLAAAGILLVVALK AALLPLVFMLEPRAYATASAPAAALFATMTVGITSLIRYTVLVFGDAAGELLVVALK LMLPLATLIGAGTALAPLQOBLIVVVASGTLVAGVLSERLAAASLTYLLH STWINGGFLIADLVAYORDKRAKDQBLVAVASGTLVAGVLSERLAAASLTYLLH FGKWLQSVAPSGQALALMVSVLGSLVALVALSRAGSTLFWPRTGHVYLSAELEHDHG
		Query Match Best Local Similarity 8.2%; Score 39.6; DB 1; Length 13946; Matches 75; Conservativity 56.0%; Pred. No. 27; Mismatches 0; Mismatches 59; Indels 0; Gaps 0;
		Qy 86 TTGGCGGGGCCACCGTGGCGGTTGTGGCGGCTGTGCCGCGACCTGTGCTGAAGCT 145 Db 7494 TCGGTTCCATCGCGCGGTGGCGGCTGTGCCGCGACCGCGCGTGGCGGCGCA 7553 Qy 146 CTGCTGCTGTGTCGATGCCATCGCGGCTGTGGCGGCTGTGCTGCTATGCTCTAAA 205 Db 7554 AGGCGACGCTACCTCGCGGCGCATCATCGCGCGGTGCGTGGCGGCTGTGCGGCTA 7613 Qy 206 TCGGTCAGAGCATG 219 Db 7614 CCGAAGAAGCGCTG 7627
		RESULT 13 ATF7K2 LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM
		ATF7K2 Arabidopsis thaliana DNA chromosome 4, BAC clone F7K2 (ESSA project). AL033545 AL033545.2 GI:5738368 Thale cress. Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. Bevan, M., Wedler, H., Wambutt, R., Bancroft, I., Mewes, H. W., Lemcke, K. and Mayer, K. F. X. Unpublished 2 (bases 92830 to 92989) Bevan, M., Volckaert, G., Grymonprez, B., Voelt, M., Robben, J., Bancroft, I., Mewes, H. W., Lemcke, K. and Mayer, K. F. X. Unpublished 3 (bases 1 to 16021) Mewes, H. W., Robben, J., Grymonprez, B., Volckaert, G., Bancroft, I., Mewes, H. W., Lemcke, K. and Mayer, K. F. X. Unpublished 4 (bases 1 to 106702) EU Arabidopsis sequencing project. Submitted (09-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@jrc.ac.uk On Aug 18, 1999 this sequence version replaced g1:3892698. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ . Location/Qualifiers 1..106702 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" 1..15769 /note="basepairs 1-15769 are not integral part of BAC clone F7K2" /complement(join(<3..98,594..821)) /gene="AT4g22380" /note="strong similarity to NHP2/RS6 FAMILY PROTEIN, Homo sapiens, PID:g4826860 contains Ribosomal protein L7Ae signature AA73-90 contains EST gb:FI3981" /codon_start=1 /product="Ribosomal protein L7Ae-like (fragment)"
		FEATURES source misc-feature CDS

[illegible]

/number-2

Query Match 8.2%; Score 39.4; DB 8; Length 106702;
Best Local Similarity 59.3%; Pred. No. 31;
Matches 67; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 85 GTTGGCGCGCCACCGGTGCGTGTGGGGGGTGGCCGCGCAGCTGTGGTAAGC 144
Db 33426 GTTGGAGGCTGCTCCGGTGTGTTGAGAGAGTCTGCGTGTGGAGGTGC 33485
QY 145 TCTGGTGTGTCGTGATGAGCCATCGCGGCGTGTCTGGCGGCGTCTATTGG 197
Db 33486 TCTGGTGTGTCGTGATGAGAGAGTCTCCGCGTGTGTTGAGAGGTCTTCG 33538

RESULT 14
LOCUS ATCHRIV57
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 57.
ACCESSION AL161557
VERSION AL161557.2 GI:7269071
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
Medler, H., Wandutt, R., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
Unpublished
2 (bases 141638 to 141797)
Volckaert, G., Grymonprez, B., Voelt, M., Robben, J., Mewes, H.W.,
Lemcke, K. and Mayer, K.F.X.
Unpublished
3 (bases 143670 to 143957)
Hilbert, H., Braun, M., Holzer, E., Brandt, A., Duesterhoeft, A.,
Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
Unpublished
4 (bases 48809 to 64829)
Robben, J., Grymonprez, B., Volckaert, G., Mewes, H.W., Lemcke, K. and
Mayer, K.F.X.
Unpublished
5 (bases 1 to 199577)
EU Arabidopsis sequencing project.
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
leemcke@mpi-biochem.mpg.de, mayer@mpi-biochem.mpg.de, Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@jbsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
this fragment has an overlap with ATCHRIV56 at the 3' end and an
overlap with ATCHRIV58 at the 3' end.

COMMENT
FEATURES
source
gene
exon
gene
CDS

1 199577
/organism="Arabidopsis thaliana"
/varietal="Colombia"
/db_xref="taxon:3702"
/chromosome="4"
6574..9057
/gene="AT4g22260"
complement(6674..6882)
/number=1
/number=1
complement(join(6674..6882,7268..7339,7507..7575,
/gene="AT4g22260".7978,8361..8454,8588..8777,8909..9057))
7667..7775,7863..7978,8361..8454,8588..8777,8909..9057))
7667..7775,7863..7978,8361..8454,8588..8777,8909..9057))
/gene="AT4g22260"

/note="similarity to alternative oxidase, Mangifera
indica, PIR2:S45035
contains EST gb:742793, A1994896.1, A1997645.1, R30022,
AA395166"
/codon_start=1
/product="putative protein"
/protein_id="CAB79181.1"
/db_xref="gi:7269072"
/translation="MAAISGSGTLLTISPLVTLRRRAAVSYSSRHLLHPLSS
KLEGVNVEFLDSVIRILDTLYKVERESTKAETSTGTEPLEPMNMSSTSTSAFETWIT
RADYIKVAFESNMENHLLIMEDLGNMWDPLAOHIAFYEPMTVLYLISPRM
AVHSECEVSHAVETDKELIKSGBELNMPADIAVYKYGGLYFENLXDPVNI
RDDEAFHCKTMRACOTLGSRSPHSLIEDDDTEERSGVPEEAHCBGIVDCKKSIIT
S"
complement(6883..7267)
/gene="AT4g22260"
/number=1
complement(7268..7339)
/gene="AT4g22260"
/number=2
complement(7340..7506)
/gene="AT4g22260"
/number=2
complement(7507..7575)
/gene="AT4g22260"
/number=3
complement(7576..7666)
/gene="AT4g22260"
/number=3
complement(7667..7775)
/gene="AT4g22260"
/number=4
complement(7776..7862)
/gene="AT4g22260"
/number=4
complement(7863..7978)
/gene="AT4g22260"
/number=5
complement(7979..8360)
/gene="AT4g22260"
/number=5
complement(8361..8454)
/gene="AT4g22260"
/number=6
complement(8455..8587)
/gene="AT4g22260"
/number=6
complement(8588..8777)
/gene="AT4g22260"
/number=7
complement(8778..8908)
/gene="AT4g22260"
/number=7
complement(8909..9057)
/gene="AT4g22260"
/number=8
10103..10244
/gene="AT4g22270"
/number=1
10103..12393
/gene="AT4g22270"
join(10103..10244,10395..10516,11327..11779,11868..12174,
12254..12393)
/note="contains EST gb:A1997575.1"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB79182.1"
/db_xref="gi:7269073"
/translation="MATNGSSSEHLPEPTPSAIIINKOPEPPASRETFNSLVLMFDO
TLEATYRIWTIISGFNOIPIIINILSHVACTLQSSMLYRSLFIIVICILVKITCH

10TTLRLDDEFAFCASEITPDVRSALGCHOKIRRNRLRIVSHRFRREILLSLITVATQFM
ALLTTTRASAVANITVEVGEIATCSLIVTGVFICLRSAATKIRHKQSVTLAKMNVIC
ATVDSFDHDEGTPTGSIETESOVSLRGNAIETSDDEEGDDDLDTKTHPIYANTIS
TQKRALVITLENKAGITVYGFVLDVRSWLNITFGIELLLMLLTKTIGILA"
10245..10394
/gene="AT4g22270"
/number=1
10395..10516
/gene="AT4g22270"
/number=2
10517..11326
/gene="AT4g22270"
/number=2
11168..11253
/gene="AT4g22270"
/note="Tyr-CRNA_(GTA)"
/product="tRNA-Tyr"
11327..11779
/gene="AT4g22270"
/number=3
11780..11867
/gene="AT4g22270"
/number=3
11868..12174
/gene="AT4g22270"
/number=4
12175..12253
/gene="AT4g22270"
/number=4
12254..12393
/gene="AT4g22270"
/number=5
12761..15111
/gene="AT4g22280"
/complement(12761..13015)
/gene="AT4g22280"
/number=1
complement(join(12761..13015,13081..13227,13379..14157,
14209..14314,14995..15111))
/gene="AT4g22280"
complement(join(12761..13015,13081..13227,13379..14157,
14209..14314,14995..15111))
/gene="AT4g22280"
/note="similarity to hypothetical protein, chr.4,
Arabidopsis thaliana, PITCHX:G2392763
contains EST gb:A198861.1, A1993240.1"
/codon_start=1
/product="putative protein"
/protein_id="CA79183.1"
/db_xref="GI:7269074"
/translation="MEKVMTECTICVGRNPATIVNAGLINSHOCMISLAKCYIIE
FQVADGLLEFLRITPERFIRIMQDRINEATSVLSMRNRYLFAFPNCLDQOE
VGGDSFIDFVRLVATGNFPIRISICRMSIDGHTRMVADLEHGVSLDIDI
ISEDIGFVLEIFCTKLEKLAGFYAMVDVLSKTLFLSSIMPCRDCCP
LGRLLSACVLEELITIGGCMQHEICRFTVSSSLKILITSSPHYEYGITIDPSTL
ATLEYSDIVRHRPVPVNLSEIVAKIALDICSNSPTNLIKLRNVEYELISGYNSO
IITDFREALIPVLSKLFRLSIKSNSRRTYIWNLLPILLESKPLETIVTIGPISADRIE
REYGLSCPVKLEITEYGGKYELEQMEHFLKLPCLLELVKVRASAINDEXSRITKD
LIMVPRSSNCNKLKFC"
complement(13016..13080)
/gene="AT4g22280"
/number=1
complement(13081..13227)
/gene="AT4g22280"
/number=2
complement(13228..13378)
/gene="AT4g22280"
/number=2
complement(13379..14157)
/gene="AT4g22280"
/number=3

Query Match

8.2%; Score 39.4; DB 8; Length 199577;

Best Local Similarity 59.3%; Pred. No. 32;
Matches 67; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
QY 85 GTTGGCGCGCCACCGGTGGTGTGGCGGTTGTCGCCGACGTTGCGTAAAGC 144
Db 82234 GTTGGAGGTGTCTCCGGTGTGTGTGGAGGATCTCTGTCGTGGAAAGTGTG 82293
QY 145 TCTGTCGTCTGTCGATGCGCATCGCGGTTCTGGCGGTCGTGATTGG 197
Db 82294 TCTGTCGTCTGTTGGAGGAGTCCCGGTGTGTGGAGGTCTCTG 82346

RESULT 15
DMU60844/c 1070 bp DNA linear INV 04-JUN-1998
LOCUS DMU60844
DEFINITION Drosophila melanogaster strain DPF-46 triose phosphate isomerase
(Tpi) gene, complete cds.
ACCESSION U60844
VERSION U60844.1 GI:3184329
KEYWORDS
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS Hasson, E., Wang, I.N., Zeng, L.W., Kreitman, M. and Eanes, W.F.
TITLE Nucleotide variation in the triosephosphate isomerase (Tpi) locus
of Drosophila melanogaster and Drosophila simulans
Mol. Biol. Evol. 15 (6), 756-769 (1998)
JOURNAL 98277691
MEDLINE
REFERENCE
AUTHORS Hasson, E., Wang, I.W., Zeng, L.W., Kreitman, M. and Eanes, W.F.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-1996) Ecology & Evolution, State University of
New York at Stony Brook, Stony Brook, NY 11794, USA

FEATURES
source
1..1070
/organism="Drosophila melanogaster"
/strain="DPF-46"
/db_xref="taxon:7227"
/chromosome="3R"
/map="99 E"
/note="individual isolate from strain DPF-46 (New York,
USA); Tpi slow allele."
join(<134..670,733..>939)
/gene="Tpi"
/product="triose phosphate isomerase"
<134..>939
/gene="Tpi"
join(134..670,733..939)
/gene="Tpi"
/codon_start=1
/product="triose phosphate isomerase"
/protein_id="AAC39049.1"
/db_xref="GI:3184330"
/translation="MSRRCVGGNKMNDQSIATLTSALDPTEYVIGCPA
ITLMYARNLPEELGAGONAVKAVAGATGELISPMALDIDGADWVILGHSEKRALEG
ESDALIAEKRAHLAELGLVIACIGETLEEREGKNEVAVRQMCAYAKIKDKMNV
VAVEPVAIGTGKTATPDDAOEVAHFRLQMLSDNISKEVSASIRLIGSGSVTAANAKE
LAKKPIDDFGVGASIKPEFDIIVARQ"

BASE COUNT 238 a 345 c 301 g 186 t
ORIGIN

Query Match 8.1%; Score 39; DB 3; Length 1070;
Best Local Similarity 46.5%; Pred. No. 36;
Matches 126; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
QY 52 GTGGGTGGCCCGCAGAACTTACAGCCGACAGAGTGGCGCGCACCGGTGGTGTG 111
Db 927 GTTGATGATGTCTCCAGCAACTCCAGGCTTCAGGAGGCGCCCTCCAGCAAGCAATCGAT 868
QY 112 GCGGCTTCCGCGCGCGTCTCGTGAAGGCTCTGTCGTGTCGATGGCATCGGC 171

us-09-677-374-3.rge

Page 18

Search completed: October 27, 2002, 17:46:41
Job time : 1387.41 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 20:15:49 : Search time 967.668 Seconds
(without alignments)
6736.842 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483
Sequence: 1 atgcgtgctgctgctgcagcggg.....agggtgattagaccacgaaaa 483

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estlin:*
4: em_estlcn:*
5: em_estlcn:*
6: em_estlcn:*
7: em_estlcn:*
8: em_hlc:*
9: gb_estl:*
10: gb_estl:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vtc:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	4.1	516	12	AQ518147 HS_5090_A
2	20	4.1	823	10	BI186311 UNL-P-FN-
3	19	3.9	190	12	AL003232 F_rubrip
4	19	3.9	348	10	TS1729 YB28d01.r1
5	19	3.9	459	9	AV763598 AV763598
6	19	3.9	527	10	BM381932 MEST542-A
7	19	3.9	528	9	AM658421
8	19	3.9	536	9	AI637152 603001D08
9	19	3.9	630	10	BM35231 MEST147-F
10	19	3.9	648	12	BM381604 AG-ND-132
11	19	3.9	697	10	BM350333 MEST264-C
12	19	3.9	744	10	BI764273 6030045965
13	19	3.9	2405	10	BM3815056 602673281
14	18	3.7	372	12	BM395820 AG-ND-119
15	18	3.7	386	9	AM145423 ga20h03.y
16	18	3.7	387	12	AZ480429 LM0301N22
17	18	3.7	393	10	BF656950 OV2_19_B0

18	18	3.7	394	12	AQ408806	AQ408806 HS_5115_A
19	18	3.7	405	10	BE552264	BE552264 hy05c03.x
20	18	3.7	405	12	B76017	B76017 RPT11-13EL
21	18	3.7	406	9	AI953147	AI953147 wq50h02.x
22	18	3.7	411	9	AI635553	AI635553 ts95e09.x
23	18	3.7	441	10	R98050	R98050 yr30d06.s1
24	18	3.7	445	9	BB748581	BB748581 BR748581
25	18	3.7	487	10	BI479759	BI479759 WHE3451.G
26	18	3.7	488	9	AI167944	AI167944 o009b08.x
27	18	3.7	509	10	BI326875	BI326875 AK070H081
28	18	3.7	516	10	BE405086	BE405086 WHE1215_E
29	18	3.7	523	9	AI628427	AI628427 ly76d03.x
30	18	3.7	524	12	BM371300	BM371300 AG-ND-131
31	18	3.7	533	9	AA864310	AA864310 am15c05.s
32	18	3.7	560	12	A2913217	A2913217 RPT1-24-1
33	18	3.7	564	12	AQ378208	AQ378208 RPT1-23-3
34	18	3.7	567	10	BI683491	BI683491 464898 MA
35	18	3.7	568	12	BM378459	BM378459 AG-ND-165
36	18	3.7	569	12	BM393470	BM393470 AG-ND-145
37	18	3.7	573	12	BM382024	BM382024 AG-ND-140
38	18	3.7	586	10	BM3712909	BM3712909 P911n-PK0
39	18	3.7	587	12	BM394407	BM394407 AG-ND-163
40	18	3.7	600	10	T67705	T67705 yc27c07.s1
41	18	3.7	601	9	AI742638	AI742638 wq44d05.x
42	18	3.7	602	12	BM072402	BM072402 RPT1-24-2
43	18	3.7	603	9	AV596360	AV596360 AV596360
44	18	3.7	616	9	AM950089	AM950089 EST362279
45	18	3.7	635	12	BM372181	BM372181 AG-ND-167

ALIGNMENTS

RESULT 1
AQ518147
DEFINITION
HS_5090_A1.F08.T7A RPT1-11 Human Male BAC Library Homo sapiens
genomic clone Plate=666 COL=15 Row=K, DNA sequence.
516 bp DNA linear GSS 05-MAY-1999

ACCSSION
AQ518147
VERSION
AQ518147.1 GI:4743329

KEYWORDS
GSS.

SOURCE
human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
I (bases 1 to 516)

AUTHORS
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

JOURNAL
MEDLINE
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

COMMENT
Library availability, please contact Pieter de Jong
libr@redj.org.med.buffalo.edu. Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 666 row: K column: 15
Seq primer: T7
Class: BAC ends
High quality sequence stop: 516.
Location/Qualifiers
1. 516
/organism="Homo sapiens"

FEATURES

SOURCE

/db_xref="taxon:9606"
 /clone_id="Plate-666 Col-15 Row-K"
 /sex="male"
 /note="Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRI Methylase. Size selected DNA was cloned into the
 pBACE3.6 vector at EcoRI sites"

BASE COUNT 122 a 156 c 129 g 104 t 5 others

Query Match
 Best Local Similarity 4.1%; Score 20; DB 12; Length 516;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 GCCTGAGGCGAGCTCTCTG 30
 Db 191 GCCTGAGGCGAGCTCTCTG 210

RESULT 2
 B1186311/c 823 bp mRNA linear EST 10-JUL-2001
 LOCUS UNL-P-FN-cf-b-07-0-UNL.s2 UNL-P-FN Sus scrofa cDNA clone
 DEFINITION B1186311
 ACCESSION B1186311 GI:14660720
 VERSION B1186311.1
 KEYWORDS EST.
 SOURCE Pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 823)
 Caetano, A.R., Johnson, R.K. and Pomp, D.
 Generation and sequence characterization of a normalized cDNA
 library from swine ovarian follicles
 Unpublished (2001)
 CONTACT: Pomp, D
 Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: dpomp@unl.edu
 Oligo-dt track not found. Not 1 site shown in beginning of sequence
 is likely internal to the message.
 Seq primer: M13 -29
 POLYA-No.

JOURNAL COMMENT
 FEATURES
 source Location/Qualifiers
 1. 823
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection
 lines"
 /db_xref="taxon:9823"
 /clone_id="UNL-P-FN-cf-b-07-0-UNL"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pTR3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The UNL-P-FN
 library is a normalized library representing porcine
 ovarian follicles, ranging between 2.0 to 10.0 mm in
 diameter, collected during 7 days of the follicular phase
 of the pig estrous cycle. This library was derived from
 the library UNL-P-F2. The tag is a string of 5-6
 nucleotides present between the Not I site and the
 oligo-dt track. The library was constructed as described
 by Bonaldi, Lennon and Soares, Genome Research 6: 791-806
 1996.

BASE COUNT 151 a 272 c 231 g 167 t 2 others
 ORIGIN TAG-SEQ=None found"

Query Match
 Best Local Similarity 4.1%; Score 20; DB 10; Length 823;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 GTTCTGCGAGGCGAGCTCT 27
 Db 730 GTTCTGCGAGGCGAGCTCT 711

RESULT 3
 FR0011976
 LOCUS F.rubripes GSS sequence, clone 072L05aE5, genomic survey sequence.
 DEFINITION AL003232.1 GI:2448802
 ACCESSION AL003232.1
 VERSION
 KEYWORDS GSS: genome survey sequence.
 SOURCE Takifugu rubripes.
 ORGANISM Takifugu rubripes.

REFERENCE 1 (bases 1 to 190)
 Edgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrانيا, Y.,
 Williams, G. and Brenner, S.
 Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource
 Centre Hinxton, Cambridge, CB10 1SB. Email: blodelphgmp.mrc.ac.uk
 V-type: plasmid
 PRIMER: M13
 DESC: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

FEATURES
 source Location/Qualifiers
 1. 190
 /organism="Takifugu rubripes"
 /db_xref="taxon:31033"
 /clone_id="Cosmid 072L05"
 /clone="072L05aE5"

BASE COUNT 56 a 38 c 40 g 56 t
 ORIGIN

Query Match
 Best Local Similarity 3.9%; Score 19; DB 12; Length 190;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 CAGGCGAGCTCTGTGATCA 34
 Db 119 CAGGCGAGCTCTGTGATCA 137

RESULT 4
 T51729 348 bp mRNA linear EST 06-FEB-1995
 LOCUS yb28d01.r1 Stratiagene fetal spleen (#937205) Homo sapiens cDNA
 DEFINITION T51729
 ACCESSION T51729
 VERSION T51729.1 GI:653589
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 348)
 Hillier, L., Lennon, G., Becker, M., Bonaldi, M.F., Ghapelli, B.,
 Chisoe, S., Dietrich, N., Dubque, T., Favelli, A., Gish, W.,
 Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
 Schellberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags

```

/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:457"
/clone="MESF542-A12"
/clone_lib="ISUM6"
/tissue_type="mixed"
/lab_host="DH10B"
/notes="Vector: pSLiP7 (4.43 kb); Site_1: EcoRI; Site_2:
NotI. Tissue samples were collected and partially pooled
prior to RNA extraction. First-strand cDNAs were prepared
from 21 individual pools of oligo-dT selected mRNAs by
priming with 21 different NotI oligo-dT tag primers
(3'-AAGTGGAGATTCGCGCCGCGCCGNNNNNNTTTTTTTTTTTT-3').
Distinguishable 'bar code' tags, (N)6, were used for each

```

separate first-strand cDNA synthesis. Hence, these bar code tags can be used to identify the mRNA pool from which a particular cDNA clone was derived. The 'bar code' tags associated with specific tissue sources are:

ATAGCC--Germinated seeds and seedlings (1, 2, 8, 11 DAG);
 ACTGGC--Mixed mature tissues (17, 21, 38, 69, 77 DAG);
 CACAGC--Kernels (3, 5, 10, 15, 20, 25, 30, DAP);
 TAAGCC--Adventitious roots (65 DAG);
 cm, 53 and 56 DAG);
 AGTACG--Immature ears (0.2-3.0 cm, 53, 56, 59 DAG);
 TGAGCC--Husks (73 DAG);
 GACGAC--Silks;
 AATCGG--unpollinated first ears; CTAAGC--ear shanks;
 GTCGAC--etiolated seedlings; ACTGAG--callus;
 GTCGAC--Cytoloximide-treated callus; GTCGAC--anaerobic treated seedlings; GTCGAC--NAA (a-Naphthalene acetic acid) treated seedlings; GATGAC--Kinetin-treated seedlings; MAGACC--ACPC (1-aminocyclopropane-1-carboxylic acid) treated seedlings; GTCGAC--ABA (Abscissic acid)-treated seedlings; TACGGA--GA (Gibberellic acid)-treated seedlings; GCAGCA--JA (Jasmonic acid)-treated seedlings; Equal amounts of first-strand cDNA from each reaction were combined and used as template for DNA polymerase II strand synthesis. After the addition of EcoRI adaptors, and 2.0 kb were directionally cloned into the EcoRI and NotI sites of the pSP17 expression vector. Plasmid DNA isolated from the library was digested with NotI to remove empty vector clones. Linear DNAs from 5.4 to 7 kb were gel purified and ligated at low concentration to promote recircularization. Ligation products were precipitated and transformed into DH10B host cells. The complexity of the resulting library was 1.2 x 10⁶.

BASE COUNT 94 a 149 c 179 g 105 t
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 19; DB 10; Length 527;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 167 TCGGCGTGGCGTTCTGGG 185
 Db 365 TCGCGGTGGCGTTCTGGG 383

RESULT 7
 LOCUS AM658421
 DEFINITION 94440 MARC BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION AM658421
 VERSION AM658421.1 GI:7424247
 KEYWORDS EST.
 SOURCE COV.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.
 Smith, T.P.L., Grose, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Pertea, G., Holt, I., Karimicheva, S., Liang, F., Quackenbush, J. and Keel, J.W.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)
 2118013
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@mail.marc.usda.gov
 Single pass sequencing. Bases called and alt-trimmed with phred

v0.980904.e. Vector identified by cross-match with the -m1nscore 18 and -m1match 12 options.
 PCR primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCAGCTACGAGC
 Plate: 62 row: P column: 3
 Seq primer: ATTAGCTGACATATAG.
 Location/Qualifiers
 1..528
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: PCMV SPOR6; Site:1: XbaI; Site:2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and placentalary."
 BASE COUNT 97 a 164 c 150 g 117 t
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 19; DB 9; Length 528;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 211 CAGAGCATGAGCAGCAGCAGG 229
 Db 67 CAGAGCATGAGCAGCAGCAGG 49

RESULT 8
 LOCUS A1637152
 DEFINITION 603001D08.x1 603 - stressed root cDNA library from Wang/Bohner lab
 ACCESSION A1637152
 VERSION A1637152.1 GI:4688482
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea.
 Walbot, V.
 1 (bases 1 to 536)
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 603001 row: D column: 08.
 Location/Qualifiers
 1..536
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone_lib="603 - stressed root cDNA library from Wang/Bohner lab"
 /tissue_type="seedling"
 /dev_stage="salt stress"
 /lab_host="E. coli XL Gold"
 /note="Organ: root; Vector: pBluescriptII SK(+); XR; Seedling stressed root cDNA library from Wang/Bohner lab"
 BASE COUNT 95 a 150 c 195 g 94 t
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 19; DB 9; Length 536;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 167 TCGGCGTGGCGTTCTGGG 185
 ||||||||||||||||
 Db 329 TCGGCGTGGCGTTCTGGG 347

RESULT 9
 BM335231
 LOCUS
 DEFINITION BM335231 630 bp mRNA linear EST 16-JAN-2002
 MEST147-F07.T3 ISUM5-RN Zea mays cDNA clone MEST147-F07 3', mRNA
 sequence.
 ACCESSION BM335231
 VERSION BM335231.1 GI:18165392
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 630)
 Men.T.J., Qiu.F., Guo.L., Ashlock,D.A and Schnable,P.S.
 Expressed Sequence Tags from B73 Maize: Various stages and tissues
 including seedlings treated with a variety of hormones
 Unpublished (2001)
 Contact: Patrick S. Schnable
 Schnable Laboratory
 Iowa State University
 6405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
 Tel: 515-294-0975
 Fax: 515-294-2299
 Email: schnable@iastate.edu
 Individual basecall and confidence value were assigned using the
 phred software.
 (<http://depts.washington.edu/ventures/collablr/direct/index.htm>#b
 rt). Overall sequence quality assessment and vector trimming were
 conducted using the Lucy software (<http://www.tigr.org/soflab/>).
 Lucy parameters were set to ensure an overall trimmed quality of
 97.5% or better without any vector fragments in the chosen
 high-quality region of each sequence. Low-quality bases between the
 poly-T and the high-quality region were replaced with N's to serve
 as spacers.
 PCR Primers
 FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)
 BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)
 Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).
 FEATURES
 source
 Location/Qualifiers
 1..630
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone="MEST147-F07"
 /clone_1lb="ISUM5-RN"
 /tissue_type="mixed"
 /lab_host="DH10B"
 /note="Vector: pTR3PAC; Site_1: EcoRI; Site_2: NotI;
 Tissues: Germinated seed and seedlings (1', 2', 8', 11 DAG),
 mixed mature tissues (17', 21', 38', 69', 77 DAG), kernels
 (3', 5', 10', 15', 20', 25', 30', 36P), Adventitious roots (65 DAG
), Tassel (3-39 cm, 53 and 56 DAG), Immature ear (0.2-3.0
 cm, 53, 56, 59 DAG), Husk (73 DAG), Silk, unpollinated
 first ear, ear shank, etiolated seedlings, callus,
 cycloheximide-treated callus, Anaerobic treated seedlings
 NAA (alpha-Naphthalene acetic acid)-treated seedlings,
 Kinetin-treated seedlings, ACP
 (1-aminocyclopropane-1-carboxylic acid)-treated seedlings,
 brassinolide-treated seedlings, ABA (Abscissic acid
)-treated seedlings, GA (Gibberellic acid)-treated
 seedlings, JA (Jasmonic acid)-treated seedlings, ds-cDNA
 molecules were generated as follows: First-strand cDNA was
 prepared from oligo-dT selected mRNA by priming with a
 NotI oligo-dT primer (5'
 AACTGAGAAATTCGCGCCGAGAAATTTTATTTTATTTTATTTT). The
 resulting DNA:RNA hybrid was treated with RNase H and used

as a template for DNA Pol-I-catalyzed second strand
 synthesis. After the addition of EcoRI adaptors, the
 ds-cDNAs were digested with NotI and size-selected. The
 resulting molecules were directionally cloned into the
 EcoRI and NotI sites of the pTR3PAC vector. The library
 then went through one round of normalization to Cot value
 of 5 based on the methods of Marcello Bento Soares (Genome
 Research 6: 791-806, 1996)."

BASE COUNT 102 a 187 c 228 g 113 t
 ORIGIN

Query Match 3.9%; Score 19; DB 10; Length 630;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 167 TCGGCGTGGCGTTCTGGG 185
 ||||||||||||||||
 Db 352 TCGGCGTGGCGTTCTGGG 370

RESULT 10
 BH381604
 LOCUS
 DEFINITION BH381604 648 bp DNA linear GSS 10-DEC-2001
 AG-ND-132N21.TR ND-TAM Anopheles gambiae genomic clone AG-ND-132N21
 , DNA sequence.
 ACCESSION BH381604
 VERSION BH381604.1 GI:17327746
 KEYWORDS GSS.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
 ; Anopheles.
 1 (bases 1 to 648)
 Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
 Direct Submission of BAC-end sequences from Anopheles gambiae
 Unpublished (2001)
 Other_GSSs: AG-ND-132N21.TR
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjoftus@tigr.org
 This clone is from an A. gambiae BAC library (ND-TAM) provided by
 F.H. Collins and sequenced by The Institute for Genomic Research
 (TIGR). The BAC library was generated from A. gambiae PEST strain
 DNA. All DNA was extracted from newly hatched first instar larvae
 to minimize the inclusion of DNA from microorganisms that inhabit
 the gut. The DNA is derived from mixed sexes of larvae. The BAC
 library was constructed at Texas A&M University BAC Center
 University, College Station, Texas 77843-2123, USA using a HindIII
 partial digest.
 Seq primer: M13 For
 Class: BAC ends.
 FEATURES
 source
 Location/Qualifiers
 1..648
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-132N21"
 /clone_1lb="ND-TAM"
 /note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 168 a 132 c 114 g 234 t
 ORIGIN

Query Match 3.9%; Score 19; DB 12; Length 648;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 252 TCTGAAAGTGAAGCC 270
 ||||||||||||||||

Db 141 TCTGGAAGTGAAGCC 159

RESULT 11
BM350333
LOCUS
DEFINITION
BM350333 697 bp mRNA linear EST 16-JAN-2002
MEST264-C11.73 ISUM5-RN Zea mays cDNA clone MEST264-C11.3, mRNA
sequence.
BM350333
VERSION
BM350333.1 GI:18174945
KEYWORDS
EST.
SOURCE
Zea mays.
ORGANISM
Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 697)
Wen,T.J., Qiu,F., Guo,L., Ashlock,D.A and Schnable,P.S.
Expressed Sequence Tags from B73 Maize: various stages and tissues
including seedlings treated with a variety of hormones
Unpublished (2001)
Contact: Patrick S. Schnable
Schnable Laboratory
Iowa State University
G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
Tel: 515-294-0975
Fax: 515-294-2299
Email: schnable@iastate.edu
Individual basecall and confidence value were assigned using the
phred software.
(<http://depts.washington.edu/ventures/collabtr/direct/index.htm#b>
rt). Overall sequence quality assessment and vector trimming were
conducted using the Lucy software (<http://www.tigr.org/softlab/lucy/>).
Lucy parameters were set to ensure an overall trimmed quality of
97.5% or better without any vector fragments in the chosen
high-quality region of each sequence. Low-quality bases between the
poly-T and the high-quality region were replaced with N's to serve
as spacers.
PCR Primers
FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)
BACKWARD: primer T3 (ATT AAC CCG CAC TAA AG)
Seq primer: primer T3 (ATT AAC CCG CAC TAA AG).
Location/Qualifiers

FEATURES

source

1..697
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="MEST264-C11"
/clone_1id="ISUM5-RN"
/tissue_type="mixed"
/lab_host="DH10B"
/note="Vector: pT73PAC; Site_1: EcoRI; Site_2: NotI;
Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),
Mixed mature tissues (17, 21, 38, 69, 77 DAG), kernels
(3, 5, 10, 15, 20, 25, 30, DAP). Adventitious roots (65 DAG
, Tassel (3-39 cm, 53 and 56 DAG), Immature ear (0.2-3.0
cm, 53, 56, 59 DAG), Husk (73 DAG), Silk, unpollinated
first ear, ear shank, etiolated seedlings, callus,
Cycloheximide-treated, callus, Anaerobic treated seedlings,
kinetin-treated seedlings, acetic acid-treated seedlings,
NAA (a-Naphthalene acetic acid)-treated seedlings,
(1-aminocyclopropane-1-carboxylic acid)-treated seedlings,
Brassinolide-treated seedlings, ABA (Abscissic acid
)-treated seedlings, GA (Gibberellic acid)-treated
seedlings, JA (Jasmonic acid)-treated seedlings, ds-cDNA
molecules were generated as follows. First-strand cDNA was
prepared from oligo-dT selected mRNA by priming with a
NotI oligo-dT primer (5'
AACTGAGAATTCGGCGCGAGAAATTTTCTTTTCTTTT). The
resulting DNA:RNA hybrid was treated with RNase H and used
as a template for DNA pol-catalyzed second strand
synthesis. After the addition of EcoRI adaptors, the
ds-cDNAs were digested with NotI and size-selected. The

resulting molecules were directionally cloned into the
EcoRI and NotI sites of the pT73PAC vector. The library
then went through one round of normalization to Cot value
of 5 based on the methods of Marcelo Bento Soares (Genome
Research 6: 791-806, 1996)."
BASE COUNT 118 a 199 c 257 g 123 t

Query Match 3.9%; Score 19; DB 10; Length 697;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 TCGGCGTGTGGTCTGGC 185
|||||
Db 370 TCGGCGTGTGGTCTGGC 388

RESULT 12
B1764273 744 bp mRNA linear EST 25-SEP-2001
LOCUS
DEFINITION
603045965F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5186200 5',
mRNA sequence.
B1764273
ACCESSION
B1764273.1 GI:15755851
KEYWORDS
EST.
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 744)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rcmail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: L1AM1465 row: a column: 17
High quality sequence stop: 692.
Location/Qualifiers

1..744
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5186200"
/clone_1id="NIH_MGC_116"
/lab_host="DH10B"
/note="Organ: pooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (ecov site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH-MGC Library."

FEATURES

source

1..744
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5186200"
/clone_1id="NIH_MGC_116"
/lab_host="DH10B"
/note="Organ: pooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (ecov site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH-MGC Library."
BASE COUNT 147 a 238 c 121 t

Query Match 3.9%; Score 19; DB 10; Length 744;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 AAGTGAAGCCGCCAGG 277
|||||
Db 348 AAGTGAAGCCGCCAGG 366

RESULT 13
BG715056/c
LOCUS
DEFINITION BG715056 2405 bp mRNA linear EST 08-MAY-2001
602673281F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4795701 5',
mRNA sequence.
ACCESSION BG715056
VERSION BG715056.1 GI:13993987
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2405)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: L1AM10679 row: b column: 22
High quality sequence stop: 5.
Location/Qualifiers
1..2405
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4795701"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-xhoI (ctcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to 10^5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 554 a 1303 c 201 g 347 t
ORIGIN

Query Match 3.9%; Score 19; DB 10; Length 2405;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 CTCGCGTTGTGGCGGTCT 119
|||||
Db 2213 CTCGCGTTGTGGCGGTCT 2195

RESULT 14
BH395820
LOCUS
DEFINITION BH395820 372 bp DNA linear GSS 11-DEC-2001
AG-ND-119B5.TF ND-TAM Anopheles gambiae genomic clone AG-ND-119B5,
DNA sequence.
ACCESSION BH395820
VERSION BH395820.1 GI:17341961
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
; Anopheles.
REFERENCE 1 (bases 1 to 372)
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

TITLE
JOURNAL
COMMENT Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)
Other_GSSs: AG-ND-119B5.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: Bjloftus@ligr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.
Location/Qualifiers
1..372
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone_image="AG-ND-119B5"
/clone_lib="ND-TAM"
/note="Vector: pCIBAC1; Site_1: HindIII"

BASE COUNT 90 a 84 c 66 g 132 t
ORIGIN

Query Match 3.7%; Score 18; DB 12; Length 372;
Best Local Similarity 100.0%; Pred. No. 1,3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 253 CTGCAAAAGTGAAGGCC 270
|||||
Db 142 CTGCAAAAGTGAAGGCC 159

RESULT 15
AW145423/c
LOCUS
DEFINITION AW145423 386 bp mRNA linear EST 01-NOV-1999
ga20H03.Y1 Moss EST library PPU Physcomitrella patens cDNA clone
PEP_SOURCE_ID:PPU021606 5', mRNA sequence.
ACCESSION AW145423
VERSION AW145423.1 GI:6167159
KEYWORDS EST.
SOURCE Physcomitrella patens.
ORGANISM Physcomitrella patens.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
1 (bases 1 to 386)
Quatrano,R., Bashlades,S., Cove,D., Cumling,A., Knight,C., Clifton
S., Marra,M., Hillier,L., Page,D., Martin,J., Wyllie,T., Underwood
K., Theising,B., Allen,K., Bowers,Y., Person,B., Swaller,T.,
Stephens,M., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
Leeds/Mash U Moss EST Project
Unpublished (1999)
Contact: Ralph Quatrano
Leeds/Mash U Moss EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Libraries were constructed by Dr. Stavros Bashlades as part of the
Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)

Seq primer: -40RP from Gibco
 High quality sequence stop: 373.
 Location/Qualifiers
 1. 386

/organism="Physcomitrella patens"
 /db_xref="taxon:3218"
 /clone="PEP_SOURCE_ID:PPU021606"
 /clone_lib="Moss EST library ppy"
 /tissue_type="protonemata: 7 day old tissue
 ammonium-grown"
 /lab_host="DH10B"

/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; Construction of the cDNA library was carried out
 using Stratagene's 'UnizAP - cDNA synthesis kit'. cDNA
 was constructed using an oligo dt primer/linker that
 contains a XhoI site within it. Following ds cDNA
 synthesis, EcoRI adapters were ligated to the blunt ends
 and sample was digested with XhoI. The result is cDNA
 with an EcoRI sticky end on one side and a XhoI sticky
 end on the other. This cDNA was ligated directionally in
 UnizAP arms. The vector is designed containing the
 Bluescript sequence as well as lambda DNA and cDNA is
 cloned within this Bluescript sequence. The vector was
 then packaged using Gold gigaPackaging extracts. Library
 was grown in XlriBlue MRF cells and amplified. The library
 was excised by mass excision using Stratagene's 'Mass
 excision kit' that uses exassist as a helper phage that
 releases the Bluescript sequence and circularises it as
 single stranded plasmids that are then packaged (by helper
 phage) and secreted out of the host cell as phagemids.
 SOUR cells were transformed with phagemids and the library
 was plated out on LB-amp plates to select for
 transformants. Approximately 1,000,000 colonies were grown
 and recovered. The double stranded plasmid library was
 recovered by using Qiagen Midi prep kit. 2 micro grams of
 each library were used to transform DH10B cells by
 electroporation."

BASE COUNT 104 a 82 c 113 g 87 t
 ORIGIN

Query Match

Best Local Similarity 3.7%; Score 18; DB 9; Length 386;
 Matches 18; Conservativity 100.0%; Pred.No. 1.3e+02;

Mismatches 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 GTTGCCCTGCAGGCAGCT 25
 ||||||||||||||||
 Db 187 GTTGCCCTGCAGGCAGCT 170

Search completed: October 27, 2002, 22:36:41
 Job time : 972.668 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 20:42:44 : Search time 248.035 Seconds
(without alignments)
7028.450 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483
Sequence: 1 atgcgtggtgctgctgcaggg.....aggtgattagcaccgaaana 483

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2891707 seqs, 1804656691 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5783414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Pending_Patents_NA_New:*
1: /cgn2_6/prodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/prodata/1/pna/US05_NEW_COMB.seq:*
3: /cgn2_6/prodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/prodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/prodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/prodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/prodata/1/pna/US10_NEW_COMB.seq2:*
8: /cgn2_6/prodata/1/pna/US60_NEW_COMB.seq3:*
9: /cgn2_6/prodata/1/pna/US60_NEW_COMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22	4.6	12848	US-10-114-170-252	Sequence 252, App
2	22	4.6	38155	US-10-114-170-79	Sequence 79, Appl
3	22	4.6	39824	US-10-053-853A-876	Sequence 876, App
4	22	4.6	48423	US-10-053-853A-484	Sequence 484, App
5	22	4.6	48908	US-10-114-170-137	Sequence 137, App
6	22	4.6	58175	US-10-053-853A-137	Sequence 137, App
7	22	4.6	91772	US-10-053-853A-1	Sequence 1, Appl
8	19	3.9	453	US-09-918-995-216	Sequence 216, App
9	19	3.9	455	US-09-919-002-8733	Sequence 8733, App
10	19	3.9	847	US-10-211-818-27	Sequence 27, Appl
11	19	3.9	2270	PCT-US02-26322-47	Sequence 47, Appl
12	19	3.9	2332	US-10-219-999-30827	Sequence 30827, A
13	19	3.9	2329	US-10-108-260A-2343	Sequence 2343, Ap
14	19	3.9	2329	US-10-108-260-2343	Sequence 2343, Ap
15	18	3.7	440	US-10-221-279-4847	Sequence 4847, Ap
16	18	3.7	466	US-09-918-995-8784	Sequence 8784, Ap
17	18	3.7	643	US-10-027-632-276565	Sequence 276565,
18	18	3.7	643	US-10-027-632-276567	Sequence 276567,
19	18	3.7	644	US-10-027-632-276566	Sequence 276566,
20	18	3.7	705	US-10-027-632-173497	Sequence 173497,
21	18	3.7	705	US-10-027-632-173498	Sequence 173498,
22	18	3.7	716	US-10-027-632-22462	Sequence 22462, A
23	18	3.7	1300	PCT-US02-25766-2948	Sequence 2948, Ap
24	18	3.7	1855	US-10-027-632-261467	Sequence 261467,

C 25	18	3.7	1855	7	US-10-027-632-261468	Sequence 261468,
C 26	18	3.7	2661	9	US-60-360-039-32215	Sequence 32215, A
C 27	18	3.7	2690	9	US-60-360-039-28064	Sequence 28064, A
C 28	18	3.7	2690	9	US-60-360-039-30818	Sequence 30818, A
C 29	18	3.7	11736	5	US-09-663-870A-1	Sequence 1, Appl
30	17	3.5	123	8	US-10-182-995-11452	Sequence 11452, A
31	17	3.5	123	8	US-10-203-134-14602	Sequence 14602, A
32	17	3.5	123	8	US-10-203-136-14617	Sequence 14617, A
33	17	3.5	123	8	US-10-182-993-14306	Sequence 14306, A
34	17	3.5	123	8	US-10-182-997-10762	Sequence 10762, A
35	17	3.5	123	8	US-10-182-998-6531	Sequence 6531, Ap
36	17	3.5	123	8	US-10-203-135-14118	Sequence 14118, A
37	17	3.5	123	8	US-10-203-137-14750	Sequence 14750, A
38	17	3.5	123	8	US-10-203-138-6710	Sequence 6710, Ap
39	17	3.5	123	8	US-10-203-139-14205	Sequence 14205, A
40	17	3.5	280	5	US-09-540-210B-2218	Sequence 2218, Ap
C 41	17	3.5	359	5	US-09-721-544-23832	Sequence 23832, A
C 42	17	3.5	390	7	US-10-155-881-6701	Sequence 6701, Ap
C 43	17	3.5	451	5	US-09-918-995-22279	Sequence 22279, A
44	17	3.5	450	5	US-09-918-995-26505	Sequence 26505, A
45	17	3.5	486	8	US-10-182-995-1547	Sequence 1547, Ap

ALIGNMENTS

RESULT 1
US-10-114-170-252/c
Sequence 252, Application US/10114170
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
Perna, Valerie
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: Novel Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Plinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
/CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/114,170
FILING DATE: 01-Apr-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296,95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ. ID NO: 252:
SEQUENCE CHARACTERISTICS:
LENGTH: 12848
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 252:

US-10-114-170-252

Query Match 4.6%; Score 22; DB 8; Length 12848;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 TAATCCGGACACCGGTAAACAGC 312
Db 3851 TAATCCGGACACCGGTAAACAGC 3830
|||||

RESULT 2

US-10-114-170-79
Sequence 79, Application US/10114170
GENERAL INFORMATION:

APPLICANT: Blattner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: Novel Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 Inch, 1.44mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/114,170
FILING DATE: 01-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296, 95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 38155
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 79:

US-10-114-170-79

Query Match 4.6%; Score 22; DB 8; Length 38155;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 TAATCCGGACACCGGTAAACAGC 312
Db 31926 TAATCCGGACACCGGTAAACAGC 31947
|||||

RESULT 3

US-10-053-853A-876
Sequence 876, Application US/10053853A
GENERAL INFORMATION:

```

; TITLE OF INVENTION: Novel Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 Inch, 1.44MB storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 137:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48908
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 137:
US-10-114-170-137

Query Match          4.6%; Score 22; DB 8; Length 48908;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 TAATCCGACACCGGTAAACAGC 312
Db 15032 TAATCCGACACCGGTAAACAGC 15011

RESULT 6
US-10-053-853A-337
; Sequence 337, Application US/10053853A
; GENERAL INFORMATION:
; APPLICANT: HAYASHI, Hideo
; APPLICANT: SHINAGAWA, Hideo
; APPLICANT: MAKINO, Kozo
; APPLICANT: HAYASHI, Tetsuya
; APPLICANT: OHNISHI, Makoto
; APPLICANT: HATTORI, Masahira
; APPLICANT: KUROKAWA, Ken
; TITLE OF INVENTION: Polynucleotide molecules and polypeptides specific to Enterohemorrhagic E. coli O157:H7 and use thereof
; FILE REFERENCE: 2002-0060A/MMC/01704
; CURRENT APPLICATION NUMBER: US/10/053,853A
; PRIOR FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: JP2001-112010
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 1866
; SEQ ID NO 337
; LENGTH: 58175
; TYPE: DNA
; ORGANISM: Escherichia coli O157:H7
US-10-053-853A-337
```

```

Query Match          4.6%; Score 22; DB 8; Length 58175;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 TAATCCGACACCGGTAAACAGC 312
Db 31033 TAATCCGACACCGGTAAACAGC 31054

RESULT 7
US-10-053-853A-1/c
; Sequence 1, Application US/10053853A
; GENERAL INFORMATION:
; APPLICANT: HAYASHI, Hideo
; APPLICANT: SHINAGAWA, Hideo
; APPLICANT: MAKINO, Kozo
; APPLICANT: HAYASHI, Tetsuya
; APPLICANT: OHNISHI, Makoto
; APPLICANT: HATTORI, Masahira
; APPLICANT: KUROKAWA, Ken
; TITLE OF INVENTION: Polynucleotide molecules and polypeptides specific to Enterohemorrhagic E. coli O157:H7 and use thereof
; FILE REFERENCE: 2002-0060A/MMC/01704
; CURRENT APPLICATION NUMBER: US/10/053,853A
; PRIOR FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: JP2001-112010
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 1866
; SEQ ID NO 1
; LENGTH: 91772
; TYPE: DNA
; ORGANISM: Escherichia coli O157:H7
US-10-053-853A-1

Query Match          4.6%; Score 22; DB 8; Length 91772;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 TAATCCGACACCGGTAAACAGC 312
Db 66951 TAATCCGACACCGGTAAACAGC 66930

RESULT 8
US-09-918-995-216
; Sequence 216, Application US/09918995
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 216
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(453)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-216

Query Match          3.9%; Score 19; DB 5; Length 453;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 AAAGTGAAGCGGCGCAGC 277
```

Db

100 AAAGTGAAGCCGCCAGG 118

RESULT 9

```
US-09-919-002-8373
; Sequence 8373, Application US/09919002
; GENERAL INFORMATION:
; APPLICANT: Leeshkowitz, Dena
; APPLICANT: Liu, Jin
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; FILE REFERENCE: 20411-752CON1
; CURRENT APPLICATION NUMBER: US/09/919,002
; PRIOR FILING DATE: 2001-07-30
; PRIOR FILING DATE: FILING DATE: 1999-07-22
; PRIOR APPLICATION NUMBER: APPLICATION NUMBER: US 09/034,341
; NUMBER OF SEQ ID NOS: 13203
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8373
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(455)
; OTHER INFORMATION: n = A,T,C or G
US-09-919-002-8373
```

Query Match

```
Best Local Similarity 3.9%; Score 19; DB 5; Length 455;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 259 AAAGTGAAGCCGCCAGG 277
Db 378 AAAGTGAAGCCGCCAGG 396
```

RESULT 10

```
US-10-211-818-27
; Sequence 27, Application US/10211818
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM030C1N
; CURRENT APPLICATION NUMBER: US/10/211,818
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: 09/758,470
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; NUMBER OF SEQ ID NOS: 722
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
; LENGTH: 847
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (717)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (793)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (804)
; OTHER INFORMATION: n equals a,t,g, or c
```

FEATURE:

```
; NAME/KEY: misc_feature
; LOCATION: (812)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (833)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-211-818-27
```

Query Match

```
Best Local Similarity 3.9%; Score 19; DB 8; Length 847;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 259 AAAGTGAAGCCGCCAGG 277
Db 290 AAAGTGAAGCCGCCAGG 308
```

RESULT 11

```
PCT-US02-26322-47
; Sequence 47, Application PC/TUS0226322
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: LU, Dyrung Alina M.
; APPLICANT: SWARNAKAR, Anita
; APPLICANT: TRANG, Y. Tom
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: EMERLING, Brooke M.
; APPLICANT: FORSYTHE, Ian J.
; APPLICANT: YAO, Monique G.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: RICHARDSON, Thomas W.
; APPLICANT: BECHA, Shanya D.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: WARREN, Bridget A.
; APPLICANT: LEHR-MASON, Patricia M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LI, Joana X.
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: GIETZEN, Kimberly J.
; APPLICANT: LAL, Preeti G.
; APPLICANT: BOROMSKY, Mark L.
; APPLICANT: ISON, Craig H.
; APPLICANT: THANGAVELOU, Kavitha
; APPLICANT: XU, Yuming
; APPLICANT: LEE, Sally
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: SPRAGUE, William W.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: DING, Li
; APPLICANT: NGUYEN, Dannel B.
; APPLICANT: HONCHELL, Cynthia D.
; APPLICANT: LUO, Wen
; APPLICANT: WALIA, Nandinder K.
; APPLICANT: MARQUIS, Joseph
; APPLICANT: JACKSON, Jennifer L.
; APPLICANT: TRAN,uyen K.
; TITLE OF INVENTION: INTRACELLULAR SIGNALING MOLECULES
; FILE REFERENCE: PF-1145 PCP
; CURRENT APPLICATION NUMBER: PCT/US02/26322
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/313,245
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 60/314,751
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/316,752
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/316,847
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/322,188
```

;; PRIOR FILING DATE: 2001-09-14
;; PRIOR APPLICATION NUMBER: US 60/326,390
;; PRIOR FILING DATE: 2001-09-28
;; PRIOR APPLICATION NUMBER: US 60/328,952
;; PRIOR FILING DATE: 2001-10-12
;; PRIOR APPLICATION NUMBER: US 60/345,468
;; PRIOR FILING DATE: 2001-10-19
;; PRIOR APPLICATION NUMBER: US 60/372,499
;; PRIOR FILING DATE: 2002-04-12
;; NUMBER OF SEQ ID NOS: 90
;; SOFTWARE: PERL Program
;; SEQ ID NO 47
;; LENGTH: 2270
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; OTHER INFORMATION: Incycle ID No: 3744219CB1
PCT-US02-26322-47

Query Match 3.9%; Score 19; DB 1; Length 2270;
Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 259 AAGTGAAGCGGCCAGG 277
|||||
Db 912 AAGTGAAGCGGCCAGG 930

RESULT 12
US-10-219-999-30827

;; Sequence 30827, Application US/10219999

;; GENERAL INFORMATION:

;; APPLICANT: Cao, Yongwei

;; APPLICANT: Edgerton, Michael D

;; APPLICANT: Hinkle, Gregory J.

;; APPLICANT: Kovalic, David K.

;; APPLICANT: Liu, Jingdong

;; APPLICANT: Stein, Joshua

;; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT

;; FILE REFERENCE: 38-10(52726)C

;; CURRENT APPLICATION NUMBER: US/10/219,999

;; CURRENT FILING DATE: 2002-08-15

;; PRIOR APPLICATION NUMBER: US 60/324,109

;; PRIOR FILING DATE: 2001-09-21

;; PRIOR APPLICATION NUMBER: US 60/312,544

;; PRIOR FILING DATE: 2001-08-15

;; NUMBER OF SEQ ID NOS: 63520

;; SEQ ID NO 30827

;; LENGTH: 2322

;; TYPE: DNA

;; ORGANISM: Zea mays

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: (1571)..(2227)

;; OTHER INFORMATION:

US-10-219-999-30827

Query Match 3.9%; Score 19; DB 8; Length 2322;
Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 CAGAACTTCAGCGGCCAGG 82
|||||
Db 104 CAGAACTTCAGCGGCCAGG 122

RESULT 13
US-10-108-260A-2343

;; Sequence 2343, Application US/10108260A

;; GENERAL INFORMATION:

;; APPLICANT: HELIX RESEARCH INSTITUTE

;; TITLE OF INVENTION: Novel full length cDNA

;; FILE REFERENCE: H1-A0106

;; CURRENT APPLICATION NUMBER: US/10/108,260A

;; CURRENT FILING DATE: 2002-03-27

;; NUMBER OF SEQ ID NOS: 5458

;; SOFTWARE: PatentIn Ver. 2.1

;; SEQ ID NO 2343

;; LENGTH: 2329

;; TYPE: DNA

;; ORGANISM: Homo sapiens

US-10-108-260A-2343

Query Match 3.9%; Score 19; DB 6; Length 2329;
Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 259 AAGTGAAGCGGCCAGG 277
|||||
Db 988 AAGTGAAGCGGCCAGG 1006

RESULT 14
US-10-108-260-2343

;; Sequence 2343, Application US/10108260

;; GENERAL INFORMATION:

;; APPLICANT: HELIX RESEARCH INSTITUTE

;; TITLE OF INVENTION: Novel full length cDNA

;; FILE REFERENCE: H1-A0106

;; CURRENT APPLICATION NUMBER: US/10/108,260

;; CURRENT FILING DATE: 2002-03-27

;; NUMBER OF SEQ ID NOS: 5458

;; SOFTWARE: PatentIn Ver. 2.1

;; SEQ ID NO 2343

;; LENGTH: 2329

;; TYPE: DNA

;; ORGANISM: Homo sapiens

US-10-108-260-2343

Query Match 3.9%; Score 19; DB 8; Length 2329;
Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 259 AAGTGAAGCGGCCAGG 277
|||||
Db 988 AAGTGAAGCGGCCAGG 1006

RESULT 15
US-10-221-279-4847/C

;; Sequence 4847, Application US/10221279

;; GENERAL INFORMATION:

;; APPLICANT: Hyseq, Inc

;; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides

;; FILE REFERENCE: 21272-046

;; CURRENT APPLICATION NUMBER: US/10/221,279

;; CURRENT FILING DATE: 2002-09-06

;; PRIOR APPLICATION NUMBER: 09/574,454

;; PRIOR FILING DATE: 2000-05-19

;; PRIOR APPLICATION NUMBER: 09/519,705

;; PRIOR FILING DATE: 2000-03-07

;; NUMBER OF SEQ ID NOS: 12360

;; SOFTWARE: Custom

;; SEQ ID NO 4847

;; LENGTH: 440

;; TYPE: DNA

;; ORGANISM: Homo sapiens

US-10-221-279-4847

Query Match 3.7%; Score 18; DB 8; Length 440;
Best Local Similarity 100.0%; Pred. No. 18;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 TGCAGGCAGCTCTCTGA 31
|||||

Tue Oct 29 08:31:00 2002

us-09-677-374-3.oligo.inpu

Page 6

Db 128 TGCAGGGCAGCTCTGTGA 111

Search completed: October 28, 2002, 00:39:40
Job time : 375.035 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 20:19:49 : Search time 35.8705 Seconds
(without alignments)
3307.484 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483

Sequence: 1 atcgctgctgctgctgcaagg.....aggtgattagcacccgaaaaa 483

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	3.7	3808	2	US-08-916-917-3
2	18	3.7	3808	2	US-08-972-631-3
3	18	3.7	3808	2	US-08-972-629-3
4	18	3.7	3808	2	US-08-972-630-3
5	18	3.7	3808	2	US-08-972-211-3
6	18	3.7	3808	3	US-09-225-170-3
7	18	3.7	4411529	4	US-09-103-840A-1
8	17	3.5	754	4	US-09-541-941B-26
9	17	3.5	1384	1	US-08-553-888A-1
10	17	3.5	1465	1	US-08-553-888A-2
11	16	3.3	1566	4	US-09-032-742-1
12	16	3.3	1566	4	US-09-032-742-12
13	16	3.3	1566	4	US-09-032-742-13
14	16	3.3	1566	4	US-09-032-742-15
15	16	3.3	1566	4	US-09-032-742-16
16	16	3.3	1566	4	US-09-032-742-18
17	16	3.3	1566	4	US-09-032-742-19
18	16	3.3	1566	4	US-08-403-852D-5
19	16	3.3	1879	4	US-08-510-646B-5
20	16	3.3	1879	4	US-09-231-818-5
21	16	3.3	1896	3	US-08-501-572-6
22	16	3.3	1896	3	US-09-040-444-6
23	16	3.3	1921	1	US-08-164-839-28
24	16	3.3	1921	1	US-08-583-799-28
25	16	3.3	1998	4	US-09-232-200-68
26	16	3.3	1998	4	US-09-232-197-68
27	16	3.3	1998	4	US-09-232-201-68

c	28	16	3.3	2019	4	US-09-063-950-3	Sequence 3, Appl
	29	16	3.3	2087	4	US-09-232-191-6	Sequence 6, Appl
	30	16	3.3	2087	4	US-09-232-200-6	Sequence 6, Appl
	31	16	3.3	2087	4	US-09-232-197-6	Sequence 6, Appl
	32	16	3.3	2087	4	US-09-232-201-6	Sequence 6, Appl
	33	16	3.3	2852	4	US-09-063-950-1	Sequence 1, Appl
	34	16	3.3	3173	3	US-08-468-859A-1	Sequence 1, Appl
	35	16	3.3	3173	3	US-08-468-859A-1	Sequence 1, Appl
	36	16	3.3	3407	1	US-08-253-155A-7	Sequence 7, Appl
	37	16	3.3	3461	3	US-08-468-859A-6	Sequence 6, Appl
	38	16	3.3	3461	3	US-08-468-859A-6	Sequence 6, Appl
	39	16	3.3	3962	2	US-08-658-665-189	Sequence 189, App
	40	16	3.3	4092	2	US-08-469-537A-106	Sequence 106, App
	41	16	3.3	4695	2	US-08-231-193A-57	Sequence 57, Appl
	42	16	3.3	4695	2	US-08-486-273A-57	Sequence 57, Appl
	43	16	3.3	4695	2	US-08-940-086A-57	Sequence 57, Appl
	44	16	3.3	4695	4	US-08-940-035A-57	Sequence 57, Appl
	45	16	3.3	5302	2	US-08-658-665-190	Sequence 190, App

ALIGNMENTS

RESULT 1
US-08-916-917-3

Sequence 3, Application US/08916917
Patent No. 5856132

GENERAL INFORMATION:

APPLICANT: Stephens, Len

APPLICANT: Braselmann, Sylvia

TITLE OF INVENTION: G-BETA-GAMMA REGULATED
PHOSPHATIDYLINOSITOL-3' KINASE

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Pennie & Edmonds, LLP

STREET: 1155 Avenue of the Americas

City: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2811

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: IBM compatible

SOFTWARE: FASTSEQ for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/916,917

FILING DATE: 15-AUG-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/672,211

FILING DATE: 27-JUN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Abrams, Samuel B.

REFERENCE/DOCKET NUMBER: 8549-0006-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-493-4935

TELEFAX: 650-493-5556

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 3808 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-916-917-3

Query Match 3.7%; Score 18; DB 2; Length 3808;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 TCAGAAACAGGAATCTA 428

DB 2951 TCAGAAACAGGAATCTA 2968

RESULT 2

US-08-972-631-3
Sequence 3, Application US/08972631
Patent No. 5856133
GENERAL INFORMATION:
APPLICANT: Stephens, Len
TITLE OF INVENTION: G-BETA-GAMMA REGULATED
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/972,631
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/672,211
FILING DATE: 27-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 8549-0005-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)854-3660
TELEFAX: (415)854-3694
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3808 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-972-631-3

Query Match

Best Local Similarity 3.7%; Score 18; DB 2; Length 3808;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 TCAGAAACAGGAATCTA 428
DB 2951 TCAGAAACAGGAATCTA 2968

RESULT 3

US-08-972-629-3
Sequence 3, Application US/08972629
Patent No. 5859201
GENERAL INFORMATION:
APPLICANT: Stephens, Len
TITLE OF INVENTION: G-BETA-GAMMA REGULATED
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park

STATE: California
COUNTRY: USA
ZIP: 94025

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/972,629
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/672,211
FILING DATE: 27-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 8549-0005-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)854-3660
TELEFAX: (415)854-3694
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3808 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-972-629-3

Query Match

Best Local Similarity 3.7%; Score 18; DB 2; Length 3808;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 TCAGAAACAGGAATCTA 428
DB 2951 TCAGAAACAGGAATCTA 2968

RESULT 4

US-08-972-630-3
Sequence 3, Application US/08972630
Patent No. 5869271
GENERAL INFORMATION:
APPLICANT: Stephens, Len
TITLE OF INVENTION: G-BETA-GAMMA REGULATED
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/972,630
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/672,211
FILING DATE: 27-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 8549-0005-999

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)854-3660
TELEFAX: (415)854-3694
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3808 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-972-630-3

Query Match 3.7%; Score 18; DB 2; Length 3808;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 411 TCAGAAACAGGAATCTA 428
|||||
DB 2951 TCAGAAACAGGAATCTA 2968

RESULT 5
US-08-672-211-3
Sequence 3, Application US/08672211
Patent No. 5874273

GENERAL INFORMATION:
APPLICANT: Stephens, Len
APPLICANT: Hawkins, Phillip T.
TITLE OF INVENTION: G-BETA-GAMMA REGULATED
TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/672,211
FILING DATE: 27-JUN-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 8549-0005-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)854-3660
TELEFAX: (415)854-3694
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3808 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-672-211-3

Query Match 3.7%; Score 18; DB 2; Length 3808;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 411 TCAGAAACAGGAATCTA 428
|||||
DB 2951 TCAGAAACAGGAATCTA 2968

RESULT 6

US-09-225-170-3
Sequence 3, Application US/09225170
Patent No. 6017763

GENERAL INFORMATION:
APPLICANT: Stephens, Len
APPLICANT: Hawkins, Phillip Thomas
APPLICANT: Braselmann, Sylvia
TITLE OF INVENTION: G-BETA-GAMMA REGULATED
TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds, LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2811

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/225,170
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/916,917
FILING DATE: 15-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Abrams, Samuel B.
REGISTRATION NUMBER: 30,605
REFERENCE/DOCKET NUMBER: 8549-0006-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3808 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: Linear
US-09-225-170-3

Query Match 3.7%; Score 18; DB 3; Length 3808;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 411 TCAGAAACAGGAATCTA 428
|||||
DB 2951 TCAGAAACAGGAATCTA 2968

RESULT 7
US-09-103-840A-1/C
Sequence 1, Application US/09103840A
Patent No. 6294328

GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA

ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match

Best Local Similarity 3.7%; Score 18; DB 4; Length 4411529;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 355 CAGGACGCCGCTCAGCAG 372
DB 2210673 CAGGACGCCGCTCAGCAG 2210656

RESULT 8

US-09-541-941B-26/c
Sequence 26, Application US/09541941B
Patent No. 6309840
GENERAL INFORMATION:
APPLICANT: Wang, Jun
APPLICANT: Shaw, Pang Chui
APPLICANT: Paul, Pui-Hay
APPLICANT: Ngan, But-and Pal Ngor Karenda Ngan
TITLE OF INVENTION: POLYMERASE CHAIN REACTION - RESTRICTION FRAGMENT LENGTH POLYMORPHISM
FILE REFERENCE: 2913/52188-2A
CURRENT APPLICATION NUMBER: US/09/541,941B
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn version 3.1
SEQ ID NO 26
LENGTH: 754
TYPE: DNA
ORGANISM: Tulipa Edulis
US-09-541-941B-26

Query Match

Best Local Similarity 3.5%; Score 17; DB 4; Length 754;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 GCGTCCCTCAGCCGGA 452
DB 413 GCGTCCCTCAGCCGGA 397

RESULT 9

US-08-553-888A-1/c
Sequence 1, Application US/08553888A
Patent No. 5723293
GENERAL INFORMATION:
APPLICANT: Huang
TITLE OF INVENTION: DIAGNOSTIC METHOD AND KIT FOR DETERMINING RH BLOOD GROUP GENOTYPE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Hoffmann & Baron, LLP
STREET: 350 Jericho Turnpike
CITY: Jericho
STATE: New York
COUNTRY: USA
ZIP: 11753
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,888A
FILING DATE: 11/06/95
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: O'Dea, Sean W.
REGISTRATION NUMBER: 37690
REFERENCE/DOCKET NUMBER: 454-5

TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 822-3550
TELEFAX: (516) 822-3582
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1384 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-553-888A-1

Query Match

Best Local Similarity 3.5%; Score 17; DB 1; Length 1384;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GCTCTGATCATATATC 39
DB 635 GCTCTGATCATATATC 619

RESULT 10

US-08-553-888A-2/c
Sequence 2, Application US/08553888A
Patent No. 5723293
GENERAL INFORMATION:
APPLICANT: Huang
TITLE OF INVENTION: DIAGNOSTIC METHOD AND KIT FOR DETERMINING RH BLOOD GROUP GENOTYPE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Hoffmann & Baron, LLP
STREET: 350 Jericho Turnpike
CITY: Jericho
STATE: New York
COUNTRY: USA
ZIP: 11753
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,888A
FILING DATE: 11/06/95
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: O'Dea, Sean W.
REGISTRATION NUMBER: 37690
REFERENCE/DOCKET NUMBER: 454-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 822-3550
TELEFAX: (516) 822-3582
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1466 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-553-888A-2

Query Match
Best Local Similarity 3.5%; Score 17; DB 1; Length 1466;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GCTCTGATCATATATC 39
DB 649 GCTCTGATCATATATC 633

RESULT 11

US-09-032-742-1/c
Sequence 1, Application US/09032742
Patent No. 6255089

GENERAL INFORMATION:
APPLICANT: Teltier, Milt
APPLICANT: Herrick-Davis, Katharine
APPLICANT: Egan, Christina C.
TITLE OF INVENTION: Constitutively Activated Serotonin
TITLE OF INVENTION: Receptors
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Laurence Weinberger
STREET: 882 S. Malack Street, Suite 103
STREET: P.O. Box 1663
CITY: West Chester
STATE: PA
COUNTRY: USA
ZIP: 19380-0053
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032.742
FILING DATE: 27-FEB-1998
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Weinberger, Laurence
REGISTRATION NUMBER: 27,965
REFERENCE/DOCKET NUMBER: 3086-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 431-1703
TELEFAX: (610) 431-4181
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1566 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-032-742-1

Query Match 3.3%; Score 16; DB 4; Length 1566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 345 TTACAACAACAGGAA 360
|||||
DB 1062 TTACAACAACAGGAA 1047

RESULT 12
US-09-032-742-12/c
Sequence 12, Application US/09032742
Patent No. 6255089
GENERAL INFORMATION:
APPLICANT: Teltier, Milt
APPLICANT: Herrick-Davis, Katharine
APPLICANT: Egan, Christina C.
TITLE OF INVENTION: Constitutively Activated Serotonin
TITLE OF INVENTION: Receptors
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Laurence Weinberger
STREET: 882 S. Malack Street, Suite 103
STREET: P.O. Box 1663
CITY: West Chester
STATE: PA
COUNTRY: USA
ZIP: 19380-0053
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032.742
FILING DATE: 27-FEB-1998
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Weinberger, Laurence
REGISTRATION NUMBER: 27,965
REFERENCE/DOCKET NUMBER: 3086-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 431-1703
TELEFAX: (610) 431-4181
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1566 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-032-742-12

Query Match 3.3%; Score 16; DB 4; Length 1566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 345 TTACAACAACAGGAA 360
|||||
DB 1062 TTACAACAACAGGAA 1047

RESULT 13
US-09-032-742-13/c
Sequence 13, Application US/09032742
Patent No. 6255089
GENERAL INFORMATION:
APPLICANT: Teltier, Milt
APPLICANT: Herrick-Davis, Katharine
APPLICANT: Egan, Christina C.
TITLE OF INVENTION: Constitutively Activated Serotonin
TITLE OF INVENTION: Receptors
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Laurence Weinberger
STREET: 882 S. Malack Street, Suite 103
STREET: P.O. Box 1663
CITY: West Chester
STATE: PA
COUNTRY: USA
ZIP: 19380-0053
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032.742
FILING DATE: 27-FEB-1998
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Weinberger, Laurence
REGISTRATION NUMBER: 27,965
REFERENCE/DOCKET NUMBER: 3086-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 431-1703
TELEFAX: (610) 431-4181
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1566 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-032-742-13

Query Match 3.3%; Score 16; DB 4; Length 1566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 345 TTACACAAACAGGAA 360
|||||
Db 1062 TTACACAAACAGGAA 1047

RESULT 14
US-09-032-742-15/c
Sequence 15, Application US/09032742
Patent No. 6255089
GENERAL INFORMATION:
APPLICANT: Teitler, Milt
APPLICANT: Herrick-Davis, Katharine
TITLE OF INVENTION: Constitutively Activated Serotonin
TITLE OF INVENTION: Receptors
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSER: Laurence Weinberger
STREET: 882 S. Mallick Street, Suite 103
CITY: West Chester
STATE: PA
COUNTRY: USA
ZIP: 19380-0053
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,742
FILING DATE: 27-FEB-1998
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Weinberger, Laurence
REGISTRATION NUMBER: 27,965
REFERENCE/DOCKET NUMBER: 3086-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 431-1703
TELEFAX: (610) 431-4181
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1566 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-032-742-15

Query Match 3.3%; Score 16; DB 4; Length 1566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 345 TTACACAAACAGGAA 360
|||||
Db 1062 TTACACAAACAGGAA 1047

RESULT 15
US-09-032-742-16/c
Sequence 16, Application US/09032742
Patent No. 6255089
GENERAL INFORMATION:
APPLICANT: Teitler, Milt
APPLICANT: Herrick-Davis, Katharine
TITLE OF INVENTION: Constitutively Activated Serotonin
TITLE OF INVENTION: Receptors
NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:
ADDRESSER: Laurence Weinberger
STREET: 882 S. Mallick Street, Suite 103
CITY: West Chester
STATE: PA
COUNTRY: USA
ZIP: 19380-0053
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,742
FILING DATE: 27-FEB-1998
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Weinberger, Laurence
REGISTRATION NUMBER: 27,965
REFERENCE/DOCKET NUMBER: 3086-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 431-1703
TELEFAX: (610) 431-4181
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1566 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-032-742-16

Query Match 3.3%; Score 16; DB 4; Length 1566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 345 TTACACAAACAGGAA 360
|||||
Db 1062 TTACACAAACAGGAA 1047

Search completed: October 27, 2002, 23:22:22
Job time: 1738.87 secs

Best Local Similarity	100.0%;	Score 483;	DB 22;	Length 483;
Matches 483;	Conservative	Pred. No. 7,1e-243.		

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

4.

```

AAFB6248
ID AAFB6248 standard; DNA; 768 bp.
XX
AC AAFB6248;
XX
DT 11-JUL-2001 (first entry)
XX
DE DNA sequence of c17e2 ospa construct with N-terminal fusion partner.
XX
KW Polkiothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OSPA; salmonid rickettsial septicæmia; rickettsial disease;
KW SRB; 17E2; fusion construct; ds.
XX
OS Piscirickettsia salmonis.
XX
SYNTHETIC.
XX
Key
CDS
location/Qualifiers
1..768
/*tag= a
/partial
/product= "C17E2 OSPA with N-terminal fusion partner."
/*note= "No stop codon is given"
1..265
/*tag= b
/*note= "DNA encoding undefined N-terminal fusion partner"
286..768
/*tag= c
/*note= "Optimised OSPA construct"
misc_feature
misc_feature
misc_feature

```

PN CA2281913-A1.
XX
XX 17-MAR-2001.
PD
XX
XX PF 17-SEP-1999; 99CA-2281913.
PF
XX
XX PR 17-SEP-1999; 99CA-2281913.
PR
XX
XX PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZY/) KUZYK M A.
XX
PI Kay WW, Burian J, Kuzyk MA;
XX
XX MPI; 2001-316844/34.
DR
XX
XX P-PDSB: AAB81128.
DR
XX
XX PT Method for protecting poliklothermic fish against salmonid rickettsial
PT septicemia and other rickettsial diseases comprises administering a
PX vaccine containing the OsPa protein of Piscirickettsia salmonis
PX
XX Example 4; Fig 5; 35pp; English.

This invention relates to a method for the protection against infection
of a poliklothermic fish by the bacterial pathogen, Piscirickettsia
salmonis. The method comprises administering an immunogenic amount of a
P. salmonis specific antigen termed OsPa or an immunogenic fragment of
OsPa in the form of a vaccine. The method is used for protecting animals
particularly poliklothermic fish, against the bacterial pathogen
P. salmonis. The method is also useful for protecting against salmonid
rickettsial septicaemia (SRS) and other rickettsial diseases.
sequence represents P. salmonis OsPa DNA termed Cl7E7 optimised for
CC expression in Escherichia coli fused to DNA encoding an undefined
CC N-terminal fusion partner. The protein encoded by this fusion construct
is used in a vaccine to create an anti-OsPa antibody response.

Sequence 768 BP: 220 A. 103 C

Query match	100.00	99.99	99.98	99.97	99.96	99.95	99.94	99.93	99.92	99.91	99.90	99.89	99.88	99.87	99.86	99.85	99.84	99.83	99.82	99.81	99.80	99.79	99.78	99.77	99.76	99.75	99.74	99.73	99.72	99.71	99.70	99.69	99.68	99.67	99.66	99.65	99.64	99.63	99.62	99.61	99.60	99.59	99.58	99.57	99.56	99.55	99.54	99.53	99.52	99.51	99.50	99.49	99.48	99.47	99.46	99.45	99.44	99.43	99.42	99.41	99.40	99.39	99.38	99.37	99.36	99.35	99.34	99.33	99.32	99.31	99.30	99.29	99.28	99.27	99.26	99.25	99.24	99.23	99.22	99.21	99.20	99.19	99.18	99.17	99.16	99.15	99.14	99.13	99.12	99.11	99.10	99.09	99.08	99.07	99.06	99.05	99.04	99.03	99.02	99.01	99.00	98.99	98.98	98.97	98.96	98.95	98.94	98.93	98.92	98.91	98.90	98.89	98.88	98.87	98.86	98.85	98.84	98.83	98.82	98.81	98.80	98.79	98.78	98.77	98.76	98.75	98.74	98.73	98.72	98.71	98.70	98.69	98.68	98.67	98.66	98.65	98.64	98.63	98.62	98.61	98.60	98.59	98.58	98.57	98.56	98.55	98.54	98.53	98.52	98.51	98.50	98.49	98.48	98.47	98.46	98.45	98.44	98.43	98.42	98.41	98.40	98.39	98.38	98.37	98.36	98.35	98.34	98.33	98.32	98.31	98.30	98.29	98.28	98.27	98.26	98.25	98.24	98.23	98.22	98.21	98.20	98.19	98.18	98.17	98.16	98.15	98.14	98.13	98.12	98.11	98.10	98.09	98.08	98.07	98.06	98.05	98.04	98.03	98.02	98.01	98.00	97.99	97.98	97.97	97.96	97.95	97.94	97.93	97.92	97.91	97.90	97.89	97.88	97.87	97.86	97.85	97.84	97.83	97.82	97.81	97.80	97.79	97.78	97.77	97.76	97.75	97.74	97.73	97.72	97.71	97.70	97.69	97.68	97.67	97.66	97.65	97.64	97.63	97.62	97.61	97.60	97.59	97.58	97.57	97.56	97.55	97.54	97.53	97.52	97.51	97.50	97.49	97.48	97.47	97.46	97.45	97.44	97.43	97.42	97.41	97.40	97.39	97.38	97.37	97.36	97.35	97.34	97.33	97.32	97.31	97.30	97.29	97.28	97.27	97.26	97.25	97.24	97.23	97.22	97.21	97.20	97.19	97.18	97.17	97.16	97.15	97.14	97.13	97.12	97.11	97.10	97.09	97.08	97.07	97.06	97.05	97.04	97.03	97.02	97.01	97.00	96.99	96.98	96.97	96.96	96.95	96.94	96.93	96.92	96.91	96.90	96.89	96.88	96.87	96.86	96.85	96.84	96.83	96.82	96.81	96.80	96.79	96.78	96.77	96.76	96.75	96.74	96.73	96.72	96.71	96.70	96.69	96.68	96.67	96.66	96.65	96.64	96.63	96.62	96.61	96.60	96.59	96.58	96.57	96.56	96.55	96.54	96.53	96.52	96.51	96.50	96.49	96.48	96.47	96.46	96.45	96.44	96.43	96.42	96.41	96.40	96.39	96.38	96.37	96.36	96.35	96.34	96.33	96.32	96.31	96.
-------------	--------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-----

Best Local Similarity	100.0%;	Score 483;	DB 22;	Length 768;
Matches 483: Conservative	100.0%;	Pred. No. 7.1e+243;		

1 ATGCGTGGTTGCCCTGCACCCCAAGC	indels 0;	Gaps 0;
------------------------------	-----------	---------

[illegible]


```

286 ATGCGTGGTCCCTGCAGAGGACGCTCTGATCATATATCTCTGTTTCTGTGGGTGC 345
QY 61 GCCCAGAACTTCAGCCGCCAGAAAGTTGGGCGGCCACCGGTGCGGTGTGGGGGTGT 120
Db 346 GCCCAGAACTTCAGCCGCCAGAAAGTTGGGCGGCCACCGGTGCGGTGTGGGGGTGT 405
QY 121 GCCGCGCAGCTGTTCGTTAAAGGCTGTGTCGTGTGTCATGATGCGCATCGCGGTGCGGT 180
Db 406 GCCGCGCAGCTGTTCGTTAAAGGCTGTGTCGTGTGTCATGATGCGCATCGCGGTGCGGT 465
QY 181 CTGGGCGGCTGTGATGGCTCTAAATGCTGTCAGAGCATGACAGCAGATAAATCAAA 240
Db 466 CTGGGCGGCTGTGATGGCTCTAAATGCTGTCAGAGCATGACAGCAGATAAATCAAA 525
QY 241 CTGACACAGCTCTGTGAAAAAGTGAAGCCGCCAGGTTACTGTTGGCGTAAATCCGAC 300
Db 526 CTGACACAGCTCTGTGAAAAAGTGAAGCCGCCAGGTTACTGTTGGCGTAAATCCGAC 585
QY 301 ACCGCTACAGCTACTGTGTGAACCGGTTCCGACCTACAGCGTTACACAAACAGAA 360
Db 586 ACCGCTACAGCTACTGTGTGAACCGGTTCCGACCTACAGCGTTACACAAACAGAA 645
QY 361 CGCCGTCAGCAGTACTGCGCGAATTTACAGAGAAACCATGATCCGAGGTCAGAAACAG 420
Db 646 CGCCGTCAGCAGTACTGCGCGAATTTACAGAGAAACCATGATCCGAGGTCAGAAACAG 705
QY 421 GAAATCTACGCGACCGCGTCCCTCAGCCGAGATGGCGCGTCGAGGTGATTAGCAGAA 480
Db 706 GAAATCTACGCGACCGCGTCCCTCAGCCGAGATGGCGCGTCGAGGTGATTAGCAGAA 765
QY 481 AAA 483
Db 766 AAA 768

```

RESULT 3

AAAF86254/C
ID AAF86254 standard; DNA: 118 BP.

AC AAF86254;

DT 11-JUL-2001 (first entry)

DE PCR primer #4 used in cloning an optimisation of OspA gene.

XX Polkiothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
SRS; 17E2; fusion construct; PCR primer; ss.

OS Piscirickettsia salmonis.

XX CA281913-A1.

PD 17-MAR-2001.

PF 17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

PA (KAYW/) KAY W M.

PA (BURJ/) BURIAN J.

PA (KUZK/) KUZK M A.

PI Kay W M, Burian J, Kuzk M A.

DR WPI; 2001-316844/34.

PT Method for protecting polkiothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the OspA protein of Piscirickettsia salmonis -
XX Example 3; Fig 4B; 35pp; English.

```

CC This invention relates to a method for the protection against infection
CC of a polkiothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC OspA in the form of a vaccine. The method is used for protecting animals,
CC particularly polkiothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis OspA gene. The OspA gene is used in the method of the
CC invention.
SQ Sequence 118 BP; 21 A; 32 C; 28 G; 37 T; 0 other;
Query Match 24.4%; Score 118; DB 22; Length 118;
Best Local Similarity 100.0%; Pred. No. 1e-51;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 208 GGTACAGAGCATGGACGACGAGATAAATCAAACTGAAACGATCTGTGAAAAAGTGAAC 267
Db 118 GGTACAGAGCATGGACGACGAGATAAATCAAACTGAAACGATCTGTGAAAAAGTGAAC 59
QY 268 GCCGCGCAGGTTACTGCTGTGCGTAATCCGACACCGGTAACAGTACTCTGTGAAC 325
Db 58 GCCGCGCAGGTTACTGCTGTGCGTAATCCGACACCGGTAACAGTACTCTGTGAAC 1

```

RESULT 4
AAAF86252
ID AAF86252 standard; DNA: 110 BP.
AC AAF86252;
DT 11-JUL-2001 (first entry)
DE PCR primer #2 used in cloning an optimisation of OspA gene.
XX Polkiothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
SRS; 17E2; fusion construct; PCR primer; ss.
XX Piscirickettsia salmonis.
OS Piscirickettsia salmonis.
XX CA281913-A1.
PD 17-MAR-2001.
PF 17-SEP-1999; 99CA-2281913.
PR 17-SEP-1999; 99CA-2281913.
XX (KAYW/) KAY W M.
XX (BURJ/) BURIAN J.
XX (KUZK/) KUZK M A.
PI Kay W M, Burian J, Kuzk M A.
DR WPI; 2001-316844/34.
PT Method for protecting polkiothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the OspA protein of Piscirickettsia salmonis -
XX Example 3; Fig 4B; 35pp; English.

CC This invention relates to a method for the protection against infection
CC of a polkiothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC OspA in the form of a vaccine. The method is used for protecting animals,
CC particularly polkiothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present

CC sequence represents a PCR primer used in the cloning and optimisation of
 CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the
 CC invention.
 XX
 SO Sequence 110 BP; 13 A; 28 C; 45 G; 24 T; 0 other;

Query Match 21.5%; Score 104; DB 22; Length 110;

Best Local Similarity 100.0%; Pred. No. 2,2e-44;
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

53 TGGGTTCGCCAGAACTTACAGCCGCCAGAGAACTTGGCCGCCACCGGTGGTGGTGG 112

1 TGGGTTCGCCAGAACTTACAGCCGCCAGAGAACTTGGCCGCCACCGGTGGTGGTGG 112

113 GCGGTGTTCGCCAGCACTGTGGTAAAGGCTGTGGTGGTGGTGG 156

61 GCGGTGTTCGCCAGCACTGTGGTAAAGGCTGTGGTGGTGGTGG 104

RESULT 5

AAAF86255/c

ID AAF86255 standard; DNA; 102 BP.

XX AAF86255;

11-JUL-2001 (first entry)

PCR primer #5 used in cloning an optimisation of Ospa gene.

Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;

KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;

XX SRS; 17E2; fusion construct; PCR primer; ss.

OS Piscirickettsia salmonis.

CA2281913-A1.

17-MAR-2001.

17-SEP-1999; 99CA-2281913.

17-SEP-1999; 99CA-2281913.

(KAWW/) KAY W W.

(BUR/) BURIAN J.

(KUZ/) KUZIK M A.

Kay WW, Burian J, Kuzik MA;

WPI: 2001-316844/34.

Method for protecting poikilothermic fish against salmonid rickettsial

PT vaccine containing the Ospa protein of Piscirickettsia salmonis

PS Example 3; Fig 4B; 35pp; English.

This invention relates to a method for the protection against infection
 CC of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
 CC P. salmonis. The method comprises administering an immunogenic amount of a
 CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
 CC P. salmonis, particularly poikilothermic fish, against the bacterial pathogen
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
 CC sequence represents a PCR primer used in the cloning and optimisation of
 CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the
 CC invention.

Sequence 102 BP; 14 A; 23 C; 30 G; 35 T; 0 other;

Query Match 21.1%; Score 102; DB 22; Length 102;

Best Local Similarity 100.0%; Pred. No. 2.5e-43;

Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

334 ACCCTACGAGCTTACCAACAAAGAGAAAGCGGCTACAGCACTACTGCGGAAATTTCAGCAG 393

102 ACCCTACGAGCTTACCAACAAAGAGAAAGCGGCTACAGCACTACTGCGGAAATTTCAGCAG 393

394 AAAGCCATGATCCAGAGTCAAGAAACAGAAATCTACGGCAC 435

42 AAAGCCATGATCCAGAGTCAAGAAACAGAAATCTACGGCAC 1

RESULT 6

AAAF86253

ID AAF86253 standard; DNA; 94 BP.

XX AAF86253;

11-JUL-2001 (first entry)

PCR primer #3 used in cloning an optimisation of Ospa gene.

Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;

KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;

XX SRS; 17E2; fusion construct; PCR primer; ss.

OS Piscirickettsia salmonis.

CA2281913-A1.

17-MAR-2001.

17-SEP-1999; 99CA-2281913.

17-SEP-1999; 99CA-2281913.

(KAWW/) KAY W W.

(BUR/) BURIAN J.

(KUZ/) KUZIK M A.

Kay WW, Burian J, Kuzik MA;

WPI: 2001-316844/34.

Method for protecting poikilothermic fish against salmonid rickettsial

PT vaccine containing the Ospa protein of Piscirickettsia salmonis

PS Example 3; Fig 4B; 35pp; English.

This invention relates to a method for the protection against infection
 CC of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
 CC P. salmonis. The method comprises administering an immunogenic amount of a
 CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
 CC P. salmonis, particularly poikilothermic fish, against the bacterial pathogen
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
 CC sequence represents a PCR primer used in the cloning and optimisation of
 CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the
 CC invention.

Sequence 94 BP; 18 A; 20 C; 35 G; 21 T; 0 other;

Query Match 15.5%; Score 75; DB 22; Length 94;

Best Local Similarity 100.0%; Pred. No. 3.4e-29;

Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

158 CGATGCCATCGCGGCTGGGCTTGGCGGCTGATTTGGCTCAAAATCGTCAGACGA 217

20 CGATGCCATCGCGGCTGGGCTTGGCGGCTGATTTGGCTCAAAATCGTCAGACGA 217

218 TGGACCCAGCAGGATA 232

Db 80 TGGACGACAGATA 94

RESULT 7

ID AAF86251 standard; DNA; 111 BP.

XX AAF86251;

DT 11-JUL-2001 (first entry)

DE PCR primer #1 used in cloning an optimisation of OspA gene.

KM Polkillothermic fish: Piscirickettsia salmonis; rickettsial pathogen;

KW vaccine: OspA; salmonid rickettsial septicemia; rickettsial disease;

XX SRS; 17E2; fusion construct; PCR primer: ss.

OS Piscirickettsia salmonis.

PN CA2281913-A1.

PD 17-MAR-2001.

PF 17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

PA (KAYW/) KAY W W.

PA (BURJ/) BURIAN J.

PA (KUZV/) KUZUK M A.

PI Kay WM, Burian J, Kuzuk MA;

PS Example 3; Fig 4B; 35pp; English.

CC This invention relates to a method for the protection against infection

CC of a polkillothermic fish by the bacterial pathogen, Piscirickettsia

CC P. salmonis. The method comprises administering an immunogenic amount of a

CC P. salmonis specific antigen termed OspA, or an immunogenic fragment of a

CC OspA in the form of a vaccine. The method is used for protecting animals,

CC particularly polkillothermic fish, against the bacterial pathogen

CC P. salmonis. The method is also useful for protecting against salmonid

CC rickettsial septicemia (SRS) and other rickettsial diseases. The present

CC sequence represents a PCR primer used in the cloning and optimisation of

CC the P. salmonis OspA gene. The OspA gene is used in the method of the

CC invention.

XX Sequence 111 BP; 16 A; 32 C; 30 G; 33 T; 0 other;

SO Query Match 15.3%; Score 74; DB 22; Length 111;

Best Local Similarity 100.0%; Pred. No. 1.1e-28; Mismatches 0; Gaps 0;

Matches 74; Conservative 0; Indels 0; Gaps 0;

XX 1 ATGGGTGTCCTGACGACGACCTCTGATCATTAATCTCTGTTTCTGTTGCGTTC 60

XX 38 ATGGGTGTCCTGACGACGACGACCTCTGATCATTAATCTCTGTTTCTGTTGCGTTC 97

XX 61 GCCCAGAACTTCAG 74

XX 98 GCCCAGAACTTCAG 111

XX 11-JUL-2001 (first entry)

DE PCR primer #6 used in cloning an optimisation of OspA gene.

KM Polkillothermic fish: Piscirickettsia salmonis; rickettsial pathogen;

KW vaccine: OspA; salmonid rickettsial septicemia; rickettsial disease;

XX SRS; 17E2; fusion construct; PCR primer: ss.

OS Piscirickettsia salmonis.

PN CA2281913-A1.

PD 17-MAR-2001.

PF 17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

PA (KAYW/) KAY W W.

PA (BURJ/) BURIAN J.

PA (KUZV/) KUZUK M A.

PI Kay WM, Burian J, Kuzuk MA;

PS Example 3; Fig 4B; 35pp; English.

CC This invention relates to a method for the protection against infection

CC of a polkillothermic fish by the bacterial pathogen, Piscirickettsia

CC P. salmonis. The method comprises administering an immunogenic amount of a

CC P. salmonis specific antigen termed OspA, or an immunogenic fragment of a

CC OspA in the form of a vaccine. The method is used for protecting animals,

CC particularly polkillothermic fish, against the bacterial pathogen

CC P. salmonis. The method is also useful for protecting against salmonid

CC rickettsial septicemia (SRS) and other rickettsial diseases. The present

CC sequence represents a PCR primer used in the cloning and optimisation of

CC the P. salmonis OspA gene. The OspA gene is used in the method of the

CC invention.

XX Sequence 110 BP; 17 A; 34 C; 27 G; 32 T; 0 other;

SO Query Match 8.3%; Score 40; DB 22; Length 110;

Best Local Similarity 100.0%; Pred. No. 7.3e-11; Mismatches 0; Gaps 0;

Matches 40; Conservative 0; Indels 0; Gaps 0;

XX 444 TCAGCCGATGCGCGCTGACGAGTGTATGACCGAATAA 483

XX 76 TCAGCCGATGCGCGCTGACGAGTGTATGACCGAATAA 37

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide seq ID NO 2213.

KW Drosophila: developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI; 2001-656860/75.
 DR P-PSDB; ABB58474.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions -
 XX
 PS Claim 1; SEQ ID NO 2213; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB16176-AB16175) and the encoded proteins
 CC (AB16176-AB16175).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 3833 BP; 1094 A; 1030 C; 858 G; 851 T; 0 other;
 Query Match
 Best Local Similarity 4.1%; Score 20; DB 23; Length 3833;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 159 GATGGCCATGCGCGCGCG 178
 Db 1428 GATGGCCATGCGCGCGCG 1447
 ID AAI59154
 XX AAI59154 standard; cDNA; 2562 BP.
 AC AAI59154;
 XX
 DT 22-OCT-2001 (first entry)
 DE Human polynucleotide SEQ ID NO 1357.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 PD
 PF 26-DEC-2000; 2000WO-US34263.
 PF
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehman T, Xu C, Xue A, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 DR WPI; 2001-44253/47.
 DR P-PSDB; AAM39998.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 1357; 10078pp; English.
 CC The invention relates to human nucleic acids (AA15798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC and thrombolytic activity, chemotactic/chemokinetic activity, haemostatic
 CC assays for receptor activity, cancer diagnosis and therapy, drug screening,
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 2562 BP; 524 A; 794 C; 791 G; 450 T; 3 other;
 Query Match
 Best Local Similarity 3.9%; Score 19; DB 22; Length 2562;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 259 AAGTGGAAGCGCGCCAGG 277
 Db 1197 AAGTGGAAGCGCGCCAGG 1215
 ID AAI52010/c
 XX AAI52010 standard; DNA; 1029 BP.
 AC AAI52010;
 XX
 DT 04-SEP-2001 (first entry)
 DE Mycobacterium tuberculosis potential drug target gene SEQ ID 64.
 XX
 KW Mycobacterium tuberculosis potential drug target gene
 KW Drug target; growth; organism viability; characterisation; ds.
 KW Mycobacterium tuberculosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO200135317-A1.
 XX
 PD 17-MAY-2001.
 PD
 PF 13-NOV-2000; 2000WO-US31152.
 PF
 PR 12-NOV-1999; 99US-0165086.
 PR 12-NOV-1999; 99US-0165124.
 PR 01-FEB-2000; 2000US-0179531.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Eisenberg D, Rotstein SH, Marcotte EM;

```

XX  WPI: 2001-329193/34.
DR  P-PSDB: AAG81159.
XX
PT  Identifying nucleotide or polypeptide sequence for use as drug target,
PT  involves providing algorithm that analyzes a functional relationship
PT  between nucleotide or polypeptide sequences, and comparing the
PT  sequences -
XX
PS  Disclosure: Page 91; 207pp: English.
XX
CC  This invention relates to a method for identifying a nucleotide or
CC  polypeptide sequence that may be a drug target, or essential for growth
CC  or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC  represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC  tuberculosis proteins which are potential drug targets. The DNA and
CC  protein sequences are used to illustrate the method of the invention. The
CC  method involves providing an unknown nucleotide or polypeptide sequences,
CC  and comparing it to a number of sequences along with at least one
CC  algorithm capable of analysing a functional relationship between
CC  nucleotide and polypeptide sequences. The method is useful for
CC  characterising the function of nucleic acids and polypeptides that may be
CC  useful as a target for a drug or essential for the growth or viability of
CC  an organism.
XX
SQ  Sequence 1029 BP; 169 A; 338 C; 341 G; 181 T; 0 other:
Query Match          3.7%; Score 18; DB 22; Length 1029;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY  355 CAGGAACGCCGTCAGCAG 372
    ||||||||||||||||
DB  75 CAGGAACGCCGTCAGCAG 58

RESULT 12
AAT66951
ID  AAT66951 standard; cDNA; 1300 BP.
XX
AC  AAT66951;
XX
DT  23-NOV-1997 (first entry)
XX
DE  Asialoglycoprotein receptor L-H2 cDNA.
XX
KW  Asialoglycoprotein receptor L-H2; AGPR; autoimmune hepatitis;
KW  autoantibody; diagnosis; liver; ds.
XX
OS  Homo sapiens.
XX
FH  Key          Location/Qualifiers
FT  CDS          191..1054
    /tag=a
XX
PN  EP773289-A2.
XX
PD  14-MAY-1997.
XX
PF  20-AUG-1996; 96EP-0113349.
XX
PR  21-AUG-1995; 95JP-0212118.
XX
PA  (TOFU ) TONEN CORP.
XX
PI  Hasegawa A, Kiyosawa K, Ohue C, Tanida E, Yagi S;
PI  Yano A;
XX
DR  WPI: 1997-261316/24.
DR  P-PSDB: AAM15246.
XX
PT  Asialoglycoprotein receptor H1 and L-H2 soluble derivatives -
PT  comprise extracellular domains, optionally also with cytoplasmic

```

```

PT  domains, useful for autoimmune hepatitis diagnosis
XX
PS  Example 2; Page 35-34; 40pp: English.
XX
CC  This cDNA clone codes for human asialoglycoprotein receptor (AGPR)
CC  L-H2 (AAM15246), a membrane protein expressed specifically by
CC  hepatocytes. The appearance of autoantibodies against AGPR can be
CC  used as an indicator for autoimmune hepatitis (AIH). The cDNA
CC  clone for AGPR L-H2 was isolated from human liver cDNA by PCR
CC  amplification (see AAT66956-61). Subregions coding for the
CC  extracellular domain, optionally also with the cytoplasmic domain,
CC  can be amplified by PCR (see AAT66966-69), cloned into appropriate
CC  vectors and used for large-scale prod. of AGPR derivatives (see
CC  AAM15251-52), free of contaminating liver antigens, in bacterial
CC  (pref. E. coli) or animal (pref. mammalian) host cells. Such AGPR
CC  L-H2 derivatives can be used in a claimed method for detecting or
CC  measuring anti-AGPR antibody. AGPR H12 cDNA has also been cloned
CC  (see AAT66950).
XX
SQ  Sequence 1300 BP; 302 A; 395 C; 350 G; 253 T; 0 other:
Query Match          3.7%; Score 18; DB 18; Length 1300;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY  445 CAGCCGATGCCGCTGG 462
    ||||||||||||||||
DB  962 CAGCCGATGCCGCTGG 979

RESULT 13
AAS59780/c
ID  AAS59780 standard; DNA; 1356 BP.
XX
AC  AAS59780;
XX
DT  13-FEB-2002 (first entry)
XX
DE  Propionibacterium acnes immunogenic protein encoding DNA #275.
XX
KW  SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW  uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW  inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW  dermatological; osteopathic; neuroprotectant; ds.
XX
OS  Propionibacterium acnes.
XX
PN  WO200181581-A2.
XX
PD  01-NOV-2001.
XX
PF  20-APR-2001; 2001WO-US12865.
XX
PR  21-APR-2000; 2000US-199047P.
PR  02-JUN-2000; 2000US-208841P.
PR  07-JUL-2000; 2000US-216747P.
XX
PA  (CORI-) CORIXA CORP.
XX
PI  Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI  L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR  WPI: 2001-616774/71.
XX
PT  Propionibacterium acnes polypeptides and nucleic acids useful for
PT  vaccinating against and diagnosing infections, especially useful for
PT  treating acne vulgaris -
XX
PS  Claim 1; SEQ ID No 275; 1069pp; English.
XX
CC  Sequences AAS59506-AAS59804 represent DNA molecules encoding
CC  Propionibacterium acnes immunogenic polypeptides. The proteins and their
CC  associated DNA sequences are used in the treatment, prevention and

```



```
XX
CC This cDNA sequence codes for the p120 catalytic subunit (see
CC AAW23948) of pig G-protein regulated phosphatidylinositol-3' kinase,
CC a heterodimeric enzyme which produces the intracellular messenger
CC phosphatidylinositol (3',4',5)-triphosphate in response to activation
CC of trimeric G protein-linked receptors. This novel protein, which
CC also contains a regulatory subunit, p101 (see AAV04633), is found in
CC cells of haematopoietic origin and is involved in immune system
CC responses which cause inflammation. p120 cDNA was obtained from a
CC pig neutrophil cDNA library using a degenerate oligonucleotide
CC probe (see AAT9713) based on an isolated p120 tryptic peptide. The
CC p120 cDNA clone in pCMW3mycpl20 is deposited as ATCC 97637. The
CC invention encompasses pig and human p101 and p120 nucleotides,
CC host cell expression systems, p101 and p120 proteins (see
CC AAW23946-49), fusion proteins, polypeptides and peptides, antibodies
CC to these proteins, and transgenic animals and knockout animals.
CC Compounds which are useful for treating inflammatory response
CC disorders can be identified by screening assays using a G protein
CC activated PI3K, or a cultured host cell that expresses the p101
CC gene. Antagonists of G protein stimulated PI3K (acting through the
CC p101 subunit, especially by disrupting the interaction between the
CC p101 and p120 subunits) can be used to treat arthritis, septic
CC shock, adult respiratory distress syndrome (ARDS), pneumonia,
CC asthma, allergies, reperfusion injury, atherosclerosis, cancer and
CC Alzheimer's disease. The nucleic acids and their products can also
CC be used for diagnosis, drug screening and clinical trial monitoring
CC of inflammatory diseases.
CC
XX
SQ Sequence 3808 BP; 1010 A; 965 C; 953 G; 880 T; 0 other;
Query Match 3.7%; Score 18; DB 19; Length 3808;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 411 TCAGAAACAGGAATCTA 428
| | | | | | | | | | | | | | | |
Db 2951 TCAGAAACAGGAATCTA 2968
```

Search completed: October 27, 2002, 20:42:30
Job time : 130.349 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 18:35:09 : Search time 934.579 Seconds
(without alignments)
10815.053 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483
Sequence: 1 atgcgtgctgctgcgcaggg.....agggtgattagcaccgaaaa 483

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 segs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenBank:*

- 1: gb_ba:*
- 2: gb_hgt:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_vt:*
- 14: gb_vt:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_ph:*
- 24: em_pat:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vt:*
- 30: em_hgt_hum:*
- 31: em_hgt_inv:*
- 32: em_hgt_other:*
- 33: em_hgtgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match Length	ID	Description

C	1	22	4.6	589	8	HJRNA1956	AJ001956	HyLomecon
C	2	22	4.6	594	8	CMRNA1959	AJ001959	Chellidoni
C	3	22	4.6	596	8	CMRNA1958	AJ001958	Chellidoni
C	4	22	4.6	596	8	CMRNA1960	AJ001960	Chellidoni
C	5	22	4.6	596	8	CMRNA1960	AJ001960	Chellidoni
C	6	22	4.6	7286	8	AE005332	AE005332	Zizania t
C	7	22	4.6	10029	1	AE005325	AE005325	Escherich
C	8	22	4.6	11809	1	AE005290	AE005290	Escherich
C	9	22	4.6	13505	1	AE005441	AE005441	Escherich
C	10	22	4.6	13624	1	AE006459	AE006459	Escherich
C	11	22	4.6	103608	9	AL139161	AL139161	Human DNA
C	12	22	4.6	222605	1	AP002555	AP002555	Escherich
C	13	22	4.6	286485	1	AP002557	AP002557	Escherich
C	14	22	4.6	291136	1	AP002556	AP002556	Escherich
C	15	21	4.3	702	8	AGHGBEHA	L13954	Agoseris he
C	16	21	4.3	705	8	MIRGBEHA	L13955	Microseris
C	17	21	4.3	706	8	KGIRGBEHM	L20484	Krigia bifl
C	18	21	4.3	706	8	KGIRGBEHM	L13948	Krigia bifl
C	19	21	4.3	706	8	LAURGBEHA	L13957	Lactuca sat
C	20	21	4.3	708	8	AF422126	AF422126	Microseri
C	21	21	4.3	748	8	MCU69708	MCU69708	Microseris
C	22	21	4.3	750	8	AF386496	AF386496	Microseri
C	23	21	4.3	750	8	AF386497	AF386497	Microseri
C	24	21	4.3	750	8	AF386498	AF386498	Microseri
C	25	21	4.3	750	8	AF386499	AF386499	Microseri
C	26	21	4.3	750	8	AF386500	AF386500	Microseri
C	27	21	4.3	750	8	AF386501	AF386501	Microseri
C	28	21	4.3	750	8	AF386502	AF386502	Microseri
C	29	21	4.3	750	8	MAU69709	MAU69709	Microseri
C	30	21	4.3	750	8	AF386503	AF386503	Microseri
C	31	21	4.3	750	8	MBU69703	MBU69703	Microseris
C	32	21	4.3	750	8	MEU69705	MEU69705	Microseris
C	33	21	4.3	751	8	AF386495	AF386495	Utrouppus
C	34	21	4.3	751	8	MEU69707	MEU69707	Microseris
C	35	21	4.3	752	8	MLU69704	MLU69704	Microseris
C	36	21	4.3	754	8	AF386493	AF386493	Agoseris
C	37	21	4.3	754	8	MDU69706	MDU69706	Microseris
C	38	21	4.3	774	8	AF262890	AF262890	Lepanthes
C	39	21	4.3	777	8	AF262892	AF262892	Lepanthes
C	40	21	4.3	779	8	AF262891	AF262891	Lepanthes
C	41	21	4.3	175144	2	AC068502	AC068502	Mus muscu
C	42	20	4.1	610	8	AF244734	AF244734	Triguera
C	43	20	4.1	616	8	AF201690	AF201690	Lawsonia
C	44	20	4.1	619	8	AF229747	AF229747	Delariprea
C	45	20	4.1	643	8	AF465874	AF465874	Milleria

ALIGNMENTS

RESULT 1
HJRNA1956/c 589 bp DNA linear PLN 23-MAR-2000
LOCUS HJRNA1956 Japonica 5.8S rRNA gene.
DEFINITION HJRNA1956
ACCESSION AJ001956
VERSION AJ001956.1 GI:3392901
KEYWORDS 5.8S ribosomal RNA; 5.8S rRNA gene; internal transcribed spacer; ITS1; ITS2.

SOURCE

ORGANISM HyLomecon japonica.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales; Papaveraceae; HyLomecon.

REFERENCE

AUTHORS Biatner,F.R. and Kaderelt,J.W.
TITLE Morphological evolution and ecological diversification of the forest-dwelling poppies (Papaveraceae: Chelidonioidae) as deduced from a molecular phylogeny of the ITS region
JOURNAL Plant Syst. Evol. 219, 181-197 (1999)
REFERENCE Biatner,F.R.
AUTHORS Biatner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-OCT-1997) Biatner F.R., Dept. of Taxonomy, Institute

for Plant Genetics and Crop Plant Research, Gatersleben, D-06466,
GERMANY

FEATURES
source Location/Qualifiers

misc_feature
1..589
/organism="Hymenoclon japonica"
/db_xref="taxon:71217"

rRNA
1..225
/note="internal transcribed spacer 1"

gene
226..387
/gene="5.8S rRNA"
/product="5.8S ribosomal RNA"

misc_feature
388..589
/gene="5.8S rRNA"

BASE COUNT 125 a 169 c 166 g 129 t
ORIGIN

Query Match
Best Local Similarity 4.6%; Score 22; DB 8; Length 589;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 GCGTGCCCTCAGCCGATGACC 457
Db 370 GCGTGCCCTCAGCCGATGACC 349

RESULT 2

CMRNA1959/c

LOCUS CMRNA1959 594 bp DNA linear PLN 23-MAR-2000
DEFINITION Chelidonium majus 5.8S rRNA gene.
ACCESSION AJ001959.1 GI:3393038
VERSION 5.8S ribosomal RNA; 5.8S rRNA gene; internal transcribed spacer;
KEYWORDS ITS1; ITS2.

SOURCE
ORGANISM Chelidonium majus.
Chelidonium majus.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
Papaveraceae; Chelidonium.
1 (bases 1 to 594)
Blattner, F.R. and Kaderelt, J.W.
Morphological evolution and ecological diversification of the
forest-dwelling poppies (Papaveraceae; Chelidoniaceae) as deduced
from a molecular phylogeny of the ITS region
Plant Syst. Evol. 219, 181-197 (1999)
2 (bases 1 to 594)
Blattner, F.R.
Direct Submission
Submitted (02-OCT-1997) Blattner F.R., Dept. of Taxonomy, Institute
for Plant Genetics and Crop Plant Research, Gatersleben, D-06466,
GERMANY

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

FEATURES
source Location/Qualifiers
1..594
/organism="Chelidonium majus"
/strain="Mainz"
/sub-species="majus"
/db_xref="taxon:71251"

misc_feature
1..229
/note="internal transcribed spacer 1"

rRNA
230..391
/gene="5.8S rRNA"
/product="5.8S ribosomal RNA"

gene
230..391
/gene="5.8S rRNA"

misc_feature
392..594
/note="internal transcribed spacer 2"

BASE COUNT 123 a 168 c 173 g 130 t
ORIGIN

Query Match
Best Local Similarity 4.6%; Score 22; DB 8; Length 594;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 GCGTGCCCTCAGCCGATGACC 457
Db 373 GCGTGCCCTCAGCCGATGACC 352

RESULT 3

CMRNA1958/c

LOCUS CMRNA1958 596 bp DNA linear PLN 23-MAR-2000
DEFINITION Chelidonium majus 5.8S rRNA gene.
ACCESSION AJ001958.1 GI:3393037
VERSION 5.8S ribosomal RNA; 5.8S rRNA gene; internal transcribed spacer;
KEYWORDS ITS1; ITS2.

SOURCE
ORGANISM Chelidonium majus.
Chelidonium majus.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
Papaveraceae; Chelidonium.
1 (bases 1 to 596)
Blattner, F.R. and Kaderelt, J.W.
Morphological evolution and ecological diversification of the
forest-dwelling poppies (Papaveraceae; Chelidoniaceae) as deduced
from a molecular phylogeny of the ITS region
Plant Syst. Evol. 219, 181-197 (1999)
2 (bases 1 to 596)
Blattner, F.R.
Direct Submission
Submitted (02-OCT-1997) Blattner F.R., Dept. of Taxonomy, Institute
for Plant Genetics and Crop Plant Research, Gatersleben, D-06466,
GERMANY

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

FEATURES
source Location/Qualifiers
1..596
/organism="Chelidonium majus"
/sub-species="asiaticum Hara"
/db_xref="taxon:71251"

misc_feature
1..229
/note="internal transcribed spacer 1"

rRNA
230..391
/gene="5.8S rRNA"
/product="5.8S ribosomal RNA"

gene
230..391
/gene="5.8S rRNA"

misc_feature
392..596
/note="internal transcribed spacer 2"

BASE COUNT 124 a 169 c 171 g 132 t
ORIGIN

Query Match
Best Local Similarity 4.6%; Score 22; DB 8; Length 596;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 GCGTGCCCTCAGCCGATGACC 457
Db 375 GCGTGCCCTCAGCCGATGACC 354

RESULT 4

CMRNA1960/c

LOCUS CMRNA1960 596 bp DNA linear PLN 23-MAR-2000
DEFINITION Chelidonium majus 5.8S rRNA gene.
ACCESSION AJ001960.1 GI:3393039
VERSION 5.8S ribosomal RNA; 5.8S rRNA gene; internal transcribed spacer;
KEYWORDS ITS1; ITS2.

SOURCE
ORGANISM Chelidonium majus.
Chelidonium majus.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
Papaveraceae; Chelidonium.
1 (bases 1 to 596)
Blattner, F.R. and Kaderelt, J.W.
Morphological evolution and ecological diversification of the

REFERENCE
AUTHORS
TITLE
JOURNAL

forest-dwelling poppies (Papaveraceae: Chelidonioidae) as deduced from a molecular phylogeny of the ITS region
Plant Syst. Evol. 219, 181-197 (1999)

JOURNAL
REFERENCE 2 (bases 1 to 596)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (02-OCT-1997) Blattner F.R., Dept. of Taxonomy, Institute for Plant Genetics and Crop Plant Research, Gatersleben, D-06466, GERMANY

FEATURES
source location/Qualifiers
1. .596
/organism="Chelidonium majus"
/strain="Neustadt/W."
/sub-species="majus"
/db_xref="taxon:71251"
1. .228
/note="Internal transcribed spacer 1"
229. .390
/gene="5.8S rRNA"
/product="5.8S ribosomal RNA"
229. .390
/gene="5.8S rRNA"
391. .596
/note="Internal transcribed spacer 2"

BASE COUNT 122 a 172 c 171 g 131 t
ORIGIN

Query Match 4.6%; Score 22; DB 8; Length 596;
Best Local Similarity 100.0%; Pred. No. 0.41;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 GCGGCGCCGACGCCGATGGCC 457
Db 373 GCGGCGCCGACGCCGATGGCC 352

RESULT 5
AF169233/c 648 bp DNA linear PLN 21-NOV-1999
LOCUS zizania texana 18S ribosomal RNA gene, partial sequence; internal
transcribed spacer 1, 5.8S ribosomal RNA gene and internal
transcribed spacer 2, complete sequence; and 25S ribosomal RNA
gene, partial sequence.
AF169233
AF169233.1 GI:6456574

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
zizania texana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Zizania.

REFERENCE 1 (bases 1 to 648)
AUTHORS Kahn, A.B. and Horne, F.
TITLE The Molecular Phylogeny of Zizania (Wild Rice) and Related Grasses
JOURNAL unpublished
REFERENCE 2 (bases 1 to 648)
AUTHORS Kahn, A.B. and Horne, F.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-1999) Biology, Southwest Texas State University,
San Marcos, TX 78666, USA

FEATURES
source location/Qualifiers
1. .648
/organism="Zizania texana"
/db_xref="taxon:34180"
/country="USA"
1. .29
/product="18S ribosomal RNA"
30. .237
/note="ITS1"
/product="Internal transcribed spacer 1"
238. .404
/product="5.8S ribosomal RNA"

misc_RNA
rRNA
misc_RNA
rRNA

misc_RNA

405. .606
/note="ITS2"
/product="Internal transcribed spacer 2"
607. .>648
/product="25S ribosomal RNA"

BASE COUNT 136 a 211 c 192 g 109 t
ORIGIN

Query Match 4.6%; Score 22; DB 8; Length 648;
Best Local Similarity 100.0%; Pred. No. 0.41;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 GCGGCGCCGACGCCGATGGCC 457
Db 376 GCGGCGCCGACGCCGATGGCC 355

RESULT 6
AE005332 7286 bp DNA linear BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 2 of 3, section 1 of
DEFINITION 55.
ACCESSION AE005332 AE005174
VERSION AE005332.1 GI:12514830.
KEYWORDS
SOURCE
ORGANISM

Escherichia coli O157:H7 EDL933.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 7286)
AUTHORS Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
Grodebeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamocis, K.,
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
Welch, R.A. and Blattner, F.R.
TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL Nature 409 (6819), 529-533 (2001)
MEDLINE 21074935
PUBMED 11206551

REFERENCE 2 (bases 1 to 7286)
AUTHORS Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
Grodebeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamocis, K.,
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
Welch, R.A. and Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES
source location/Qualifiers
1. .7286
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
1. .>7286
/note="O-island #52; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
316. .642
/gene="21901"
316. .642
/gene="21901"
/gene="21901"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933X"
/protein_id="AAG55994.1"
/db_xref="GI:12514831"
/translation="MTALLLEIKANLVNDADDDMMIDKVRQATVLLVLYIGQSR"

misc_feature
gene
CDS

```
gene
CDS
        DKAISEGELIPGEALTRMKGAHMLTGMLYRNDLAERDLVQGEIPSVSLIIDL
        RCPYVL"
        /gene="21902"
        /gene="21902"
        /gene="21902"
        /function="putative enzyme; DNA packaging, phage assembly
        (phage or prophage related)"
        /note="Residues 6 to 116 of 116 are 44.14 pct identical to
        residues 4 to 112 of 112 from Genpept 118 :
        g1|6901597|gb|AAE31100.1| (AF069529) putative head-tail
        adaptor [Bacteriophage HK97]"
        /codon_start=1
        /transl_table=11
        /product="putative head-tail adaptor of prophage CP-933X"
        /protein_id="AAG55995.1"
        /db_xref="GI:12514832"
        /translation="MAISAGRLTOMISVLNLYVRNAGTEEMWVSCGKIHADIGR
        SSREMQSAEMAOAEIRIWMVNGSGRETTAASRLHLVSGFWRDRLINVLPPVPT
        GGRLEILCRIGGER"
        /gene="21903"
        /gene="21903"
        /gene="21903"
        /function="orf; Other or unknown (phage or prophage
        related)"
        /note="Residues 1 to 55 of 81 are 67.27 pct identical to
        residues 1 to 55 of 149 from Genpept 118 :
        g1|6901598|gb|AAE31101.1| (AF069529) gp10 [Bacteriophage
        HK97]"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933X"
        /protein_id="AAG55996.1"
        /db_xref="GI:12514833"
        /translation="MIETLDFSGLEDISRDQLISGAENNRVLEATRAGANVLEE
        VASXGTQCKTAPQCGGPPALPRMRGICRIPWC"
        /gene="21905"
        /gene="21905"
        /gene="21905"
        /function="orf; Other or unknown (phage or prophage
        related)"
        /note="Residues 1 to 88 of 88 are 73.86 pct identical to
        residues 62 to 149 of 149 from Genpept 118 :
        g1|6901598|gb|AAE31101.1| (AF069529) gp10 [Bacteriophage
        HK97]"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933X"
        /protein_id="AAG55997.1"
        /db_xref="GI:12514834"
        /translation="MYLSRSRSDGNEGCVHIRGVNPDGNSDNTMKADNPRNMFYWR
        FVEKGVNMPHFVRPAFDVRESEAQAVALRNRKALDEVLR"
        /gene="21906"
        /gene="21906"
        /gene="21906"
        /function="orf; Other or unknown (phage or prophage
        related)"
        /note="Residues 1 to 113 of 114 are 49.12 pct identical to
        residues 1 to 114 of 115 from Genpept 118 :
        g1|6901599|gb|AAE31102.1| (AF069529) gp11 [Bacteriophage
        HK97]"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933X"
        /protein_id="AAG55998.1"
        /db_xref="GI:12514835"
        /translation="MTADLYPYLAHLAGOVYVYVPLLDGSPVALFWVFLISS
        VSADYWGQAESSVQIDVYAGTVQARQIRQDAREAIIMLAPGSVSEMDYIPENR
        CYRALEFQYVTV"
        /gene="21908"
        /gene="21908"
        /gene="21908"
        /function="putative structure; Structural component (phage
        or prophage related)"
        /note="Residues 25 to 258 of 262 are 67.09 pct identical
        to residues 1 to 234 of 234 from Genpept 118 :
        g1|6901588|gb|AAE31091.1| (AF069529_4 (AF069529) major tail
        subunit [Bacteriophage HK97]"
        /codon_start=1
        /transl_table=11
        /product="putative tail component of prophage CP-933X"
        /protein_id="AAG55999.1"
        /db_xref="GI:12514836"
        /translation="MTSPFRSTPIPRRVRLVLSGMSALYERSQLTOYMTSSARA
        YLNOAMALDAIDNDVTAFKQVPSGKGFFLAKEVHWHSSGTCGVYAAFTSLRL
        KGPVSVYVFLAVKNEKTLTVNTGALLVMSVNGGTPPYKAKRKDQPVGQTT
        DTSKAAQSGDGAAYTCMVMDARQPSITSDACTVYVNGAG"
        /gene="21910"
        /gene="21910"
        /gene="21910"
        /function="orf; Other or unknown (phage or prophage
        related)"
        /note="Residues 1 to 77 of 80 are 62.33 pct identical to
        residues 1 to 76 of 224 from Genpept 118 :
        g1|6901601|gb|AAE31104.1| (AF069529) gp14 [Bacteriophage
        HK97]"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933X"
        /protein_id="AAG56000.1"
        /db_xref="GI:12514837"
        /translation="MAQDLKTLALRLSGFRHKTAVPEWRNVSVLREFAEWYIM
        OEVLNGDCEDDPLISVAKTRNLLEADVTILFCDSPV"
        /gene="21912"
        /gene="21912"
        /gene="21912"
        /function="orf; Other or unknown (phage or prophage
        related)"
        /note="Residues 1 to 92 of 93 are 60.86 pct identical to
        residues 133 to 224 of 224 from Genpept 118 :
        g1|6901601|gb|AAE31104.1| (AF069529) gp14 [Bacteriophage
        HK97]"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933X"
        /protein_id="AAG56001.1"
        /db_xref="GI:12514838"
        /translation="MRLALRLGRTLSLRISLSASPMAMMEFRISPLGDERGDIRN
        AOIVAVAGAGCMNVALKDMLCWDGEDEKPEVDPALEDAISLAAGS"
        /gene="21913"
        /gene="21913"
        /gene="21913"
        /function="putative structure; Structural component (phage
        or prophage related)"
        /note="Residues 1 to 1022 of 1026 are 52.47 pct identical
        to residues 1 to 1079 of 1089 from Genpept 118 :
        g1|6901589|gb|AAE31092.1| (AF069529_5 (AF069529) tail length
        tape measure protein [Bacteriophage HK97]"
        /codon_start=1
        /transl_table=11
        /product="putative tail component of prophage CP-933X"
        /protein_id="AAG56002.1"
        /db_xref="GI:12514839"
        /translation="MATLRELIRIKISNSQSFQSEIQRASRMGSEYRPTLNGGROA
        AAAREORARALAEIINDISORTGTAFSDNALFARSASMRDEYGSADDTKYRAISLG
        LKISGASRAEGSVITQPSALAGVLRGEFNSVNESGDRIVRLAAGMGYARADSK
        AMAADGKRLTRAKVVPALISQILRDREYAMPETVSSITTKVKNFPMAMVGANRAGS
```

gene

gene
CDSgene
CDSgene
CDSgene
CDS

CDS

gene
CDSgene
CDSgene
CDS

Query Match 4.6%: Score 22: DB 1: Length 7286;
 Best Local Similarity 100.0%: Pred. No. 0.47;
 Matches 22: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Qy 291 TAATCGGACACCGGTAAACAC 312
 ||||||||||||||||||
 Db 1242 TAATCGGACACCGGTAAACAC 1263

RESULT 7
 AE005325
 LOCUS
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 149
 OF 155.
 ACCESSION AE005325
 VERSION AE005325.1
 KEYWORDS GI:12514723
 SOURCE Escherichia coli O157:H7 EDL933.
 ORGANISM Escherichia coli O157:H7 EDL933.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE
 AUTHORS 1 (bases 1 to 10029)
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.
 Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
 Nature 409 (6819), 529-533 (2001)

TITLE
 JOURNAL MEDLINE
 PUBMED 21074935
 REFERENCE 11206551

AUTHORS 2 (bases 1 to 10029)
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.
 Direct Submission
 Submitted (22-OCT-2000) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

TITLE
 JOURNAL
 FEATURES
 source
 misc_feature
 gene
 CDS
 gene
 CDS

/gene="21809"
 /function="orf; Other or unknown (Phage or Prophage
 Related)"
 /note="Residues 1 to 148 of 148 are 75.16 pct identical to
 residues 1 to 149 of 149 from GenPept 118 :
 g1169015981gb|AAE31101.1| (AF069529) gp10 [Bacteriophage
 HK97]"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933N"
 /protein_id="AAG55909.1"
 /db_xref="GI:12514725"
 /translation="MTEILDFSGLEDISDRIQLLSGAENRRLREATRAGANVLKEE
 VSRARVRKRLRRNVYVLSRRSRDGMESGVIRVNPDTGSDTMTKADNRFNAFY
 WRFEVGMVMPHPVPRAPFADVRSEDAOVAIARNMRAIDVLR"
 752. .1096
 /gene="21810"
 752. .1096
 /function="orf; Other or unknown (Phage or Prophage
 Related)"
 /note="Residues 1 to 113 of 114 are 49.12 pct identical to
 residues 1 to 114 of 115 from GenPept 118 :
 g1169015991gb|AAE31102.1| (AF069529) gp11 [Bacteriophage
 HK97]"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933N"
 /protein_id="AAG55910.1"
 /db_xref="GI:12514726"
 /translation="MTEADLYPLAHAGGOVPPVYVPLIDGRPSVALPWPVPSLSS
 VSDVWGGQAESSVQIDVYACTVQARQIRDAREALMLAPGSVSEMDYIPENR
 CYRATLEFQVTV"
 1093. .1878
 /gene="21811"
 1093. .1878
 /function="putative structural component (Phage
 or Prophage Related)"
 /note="Residues 24 to 257 of 261 are 67.09 pct identical
 to residues 1 to 234 of 234 from GenPept 118 :
 g1169015881gb|AAE31091.1|AF069529_4 (AF069529) major tail
 subunit [Bacteriophage HK97]"
 /codon_start=1
 /transl_table=11
 /product="putative tail component encoded by prophage
 CP-933N"
 /protein_id="AAG55911.1"
 /db_xref="GI:12514727"
 /translation="MTFSFYKTHTPRAGFLISGRMSALYRSOLTOVMISSAPRT
 AETMDKAEYLRDCTIKVOFTAGQKODIDVTFLCSTEDENINGLGASETISMSGNT
 LNOQNALRDAYNDALYARFVLPSSGKFLAEVRQHTWSSGTNGVVAATSLRLK
 GKPFSEVPLAPFKNPDKLTVTGALLTWSVSVNGSTPPYKHAMKKDQGPVEGQTTD
 TFSKPGQASDADAKYTCVTVDSAKQSVTSVACTVTSAAAG"
 1845. .2267
 /gene="21812"
 1845. .2267
 /function="putative factor; Structural component (Phage or
 Prophage Related)"
 /note="Residues 17 to 140 of 140 are 62.90 pct identical
 to residues 1 to 123 of 224 from GenPept 118 :
 g1169016011gb|AAE31104.1| (AF069529) gp14 [Bacteriophage
 HK97]"
 /codon_start=1
 /transl_table=11
 /product="putative tail assembly chaperone encoded by
 prophage CP-933N"
 /protein_id="AAG55912.1"
 /db_xref="GI:12514728"
 /translation="MHRDSERSRRIRGWYIMKKDLTLALRLSGFRHKKVQVPEWNN
 VSVLRPSAEAWYLNQEVYNGDGDPTLSVAKTRRNLEADVTLFCDVLCOTDQGR
 VTPDDREQVLAIVGPNHARLLRQALELADASARKK"


```

/note="Residues 1 to 148 of 148 are 75.16 pct identical to
residues 1 to 149 of 149 from GenPept 118 :
g116901598|gb|AAJ31101.1| (AF069529) Gp10 [Bacteriophage
HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933M"
/protein_id="AAG5501.1"
/db_xref="GI:12514212"
/translation="MIEILLDESGLEDISRDILQLSGAENNRVLRPATRAGANVLEE
WVSRAPIRGRKMLPRNVVLSRRSGDMSGVIRGVPDGTMSDNTMTADNRMAFY
WRFVEMGTVMPPVPRPAFVRSQAQVAIAIRNRAIDEVLR"
752..1096
/gene="21369"
752..1096
/gene="21369"
/function="Orf: Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 113 of 114 are 48.24 pct identical to
residues 1 to 114 of 115 from GenPept 118 :
g116901599|gb|AAJ31102.1| (AF069529) Gp11 [Bacteriophage
HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933M"
/protein_id="AAG5502.1"
/db_xref="GI:12514213"
/translation="MTEADLYPHLAHLACGOVYVYVPLDGRPSVALPMVVFSLSS
VSADVMGGAESSVSQIIVYAGTVQARQIRQDAREXIMLLAPGSVSEMDVTPENR
CYRATLEPROTV"
1093..1872
/gene="21370"
1093..1872
/gene="21370"
/function="putative structure: Structural component (Phage
or Prophage Related)"
/note="Residues 22 to 255 of 259 are 66.66 pct identical
to residues 1 to 234 of 234 from GenPept 118 :
g116901588|gb|AAJ31091.1|AF069529_4 (AF069529) major tail
subunit [Bacteriophage HK97]"
/codon_start=1
/transl_table=11
/product="putative tail component encoded by cryptic
prophage CP-933M"
/protein_id="AAG5503.1"
/db_xref="GI:12514214"
/translation="MNEFNRTHNPPRAGFLLSGRMSALYERSOLQVMISSAPATAE
TMEKAVLRIDCTIKRVOFTAGCKODIDVTLCTROENINGIASSEISMSGNFYLN
QONALRLDAYINDNTVAFKVPQPSGKGFLEVRQHTSSGNGVAAFTSLRLKG
PVSIVPLAFVKNLDTLTVTGALLTMSVSVNGGTPPYKHAARKDGPVEGQTTDTF
SKANTSGDKGAVTCEVTSADQPOSITSDACTIVNAGAG"
1878..2252
/gene="21371"
1878..2252
/gene="21371"
/function="putative factor: Structural component (Phage or
Prophage Related)"
/note="Residues 1 to 124 of 124 are 62.90 pct identical to
residues 1 to 123 of 224 from GenPept 118 :
g116901601|gb|AAJ31104.1| (AF069529) Gp14 [Bacteriophage
HK97]"
/codon_start=1
/transl_table=11
/product="putative tail assembly chaperone encoded by
cryptic prophage CP-933M"
/protein_id="AAG5504.1"
/db_xref="GI:12514215"
/translation="MAKDLKTLAARLSEFRHKTVKVPPEMRNVSVLREPSAEAWYLM
QEVLNGDGEDDDTLTVAKTRNLLEADVTLFCVDVLCDTDLQRFVAFDDREOVLAVYGP
VHARLRQALELTLADAESARKK"
2276..2557
/gene="21372"
2276..2557
/gene="21372"
/function="Orf: Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 90 of 93 are 60.00 pct identical to
residues 133 to 222 of 224 from GenPept 118 :
g116901601|gb|AAJ31104.1| (AF069529) Gp14 [Bacteriophage
HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933M"
/protein_id="AAG5505.1"
/db_xref="GI:12514216"
/translation="MRLALRLGRTLSLRLHSLSASAFAMMMERFDRVPLGDERGDIRN
AQIVKAVFGAQCNNVALKDALMLCWDGDEDEKPEVDPAALDALSLAMS"
2828..3610
/gene="21373"
2828..3610
/gene="21373"
/function="Orf: Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 241 of 260 are 79.66 pct identical to
residues 71 to 311 of 1089 from GenPept 118 :
g116901589|gb|AAJ31092.1|AF069529_5 (AF069529) tail length
tape measure protein [Bacteriophage HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933M"
/protein_id="AAG5506.1"
/db_xref="GI:12514217"
/translation="MXGAFXGAFATGHLISLADWESSVXRLKQAXSSDEFSSXOKV
LMDISQRTXAFSDNXLFXRASAAREYGSADVLKTEAISTGKISGASTABAG
SVTQPSQALACGLRGEFNSVNESGDRIVRLAAGMVGARDLKAMDDCOLTADK
VVALISOLEVLRDEYEAAMPETVSDGITEVNAFMVWVGANFASGVTTLISVLNGV
ACQIDNVATARGTGCRRCRCPVLMQYGLASDCHGTCGCGCT"
3477..3812
/gene="21374"
3477..3812
/gene="21374"
/function="Orf: Other or unknown (Phage or Prophage
Related)"
/note="Residues 14 to 107 of 111 are 35.41 pct identical
to residues 300 to 394 of 1089 from GenPept 118 :
g116901589|gb|AAJ31092.1|AF069529_5 (AF069529) tail length
tape measure protein [Bacteriophage HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933M"
/protein_id="AAG5507.1"
/db_xref="GI:12514218"
/translation="MLPDRLLIMQOPVGLAVGAVRYFGNMAAGSAPDAGLVTAAR
NRYALAEAPFRGQITAFARARXAVYAAQQAQVAARKEEMQIAEARLARTQERLNNI
AKRTAAPQC"
4033..4464
/gene="21375"
4033..4464
/gene="21375"

```

```

CDS
/gene="21372"
2276..2557
/gene="21372"
/function="Orf: Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 90 of 93 are 60.00 pct identical to
residues 133 to 222 of 224 from GenPept 118 :
g116901601|gb|AAJ31104.1| (AF069529) Gp14 [Bacteriophage
HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933M"
/protein_id="AAG5505.1"
/db_xref="GI:12514216"
/translation="MRLALRLGRTLSLRLHSLSASAFAMMMERFDRVPLGDERGDIRN
AQIVKAVFGAQCNNVALKDALMLCWDGDEDEKPEVDPAALDALSLAMS"
2828..3610
/gene="21373"
2828..3610
/gene="21373"
/function="Orf: Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 241 of 260 are 79.66 pct identical to
residues 71 to 311 of 1089 from GenPept 118 :
g116901589|gb|AAJ31092.1|AF069529_5 (AF069529) tail length
tape measure protein [Bacteriophage HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933M"
/protein_id="AAG5506.1"
/db_xref="GI:12514217"
/translation="MXGAFXGAFATGHLISLADWESSVXRLKQAXSSDEFSSXOKV
LMDISQRTXAFSDNXLFXRASAAREYGSADVLKTEAISTGKISGASTABAG
SVTQPSQALACGLRGEFNSVNESGDRIVRLAAGMVGARDLKAMDDCOLTADK
VVALISOLEVLRDEYEAAMPETVSDGITEVNAFMVWVGANFASGVTTLISVLNGV
ACQIDNVATARGTGCRRCRCPVLMQYGLASDCHGTCGCGCT"
3477..3812
/gene="21374"
3477..3812
/gene="21374"
/function="Orf: Other or unknown (Phage or Prophage
Related)"
/note="Residues 14 to 107 of 111 are 35.41 pct identical
to residues 300 to 394 of 1089 from GenPept 118 :
g116901589|gb|AAJ31092.1|AF069529_5 (AF069529) tail length
tape measure protein [Bacteriophage HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933M"
/protein_id="AAG5507.1"
/db_xref="GI:12514218"
/translation="MLPDRLLIMQOPVGLAVGAVRYFGNMAAGSAPDAGLVTAAR
NRYALAEAPFRGQITAFARARXAVYAAQQAQVAARKEEMQIAEARLARTQERLNNI
AKRTAAPQC"
4033..4464
/gene="21375"
4033..4464
/gene="21375"

Query Match 4 6%: Score 22; DB 1: Length 11809;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 TAAATCCGACACCGGTAAACAGC 312
|||||
Db 551 TAAATCCGACACCGGTAAACAGC 572

RESULT 9
AE005441/c 13505 bp DNA 1linear BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 60
DEFINITION

```

gene

of 290.
ACCESSION AE005441 AE005174
VERSION AE005441.1 GI:12516368
KEYWORDS
SOURCE ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 13505)
Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
Postfal, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
Grobeck, E.J., Davis, N.W., Lim, A., Dimantanta, E., Potamoukis, K.,
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
Welch, R.A. and Blattner, F.R.
Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
11206551
TITLE
MEDLINE
PUBMED
REFERENCE
AUTHORS
2 (bases 1 to 13505)
Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
Postfal, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
Grobeck, E.J., Davis, N.W., Lim, A., Dimantanta, E., Potamoukis, K.,
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
Welch, R.A. and Blattner, F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1. 13505
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
<1..>13505
/note="O-island #93; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655; Cryptic prophage
CP-933v; Includes one copy of the 21 bp direct repeat that
flanks the prophage; encodes stx1a and stx1b subunits of a
Shiga-like toxin"
/complement(198..374)
/gene="23316"
complement(198..374)
/gene="23316"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded within prophage CP-933v"
/protein_id="AAG57205.1"
/db_xref="GI:12516369"
/translation="MNGSAGLVYSPYHIAIONDXNGOIGPEAGSLVQLIDRQVA
VMLSMRDGCMLSG"
complement(371..3280)
/gene="23318"
complement(371..3280)
/function="putative structure; Structural component (Phage
or Prophage Related)"
/note="Residues 1 to 962 of 969 are 53.92 pct identical to
residues 1 to 1015 of 1089 from GenPept 118 :
g116901589|gb|AAJ31092.1|AF069529.5 (AF069529) tail length
tape measure protein [Bacteriophage HK97]"
/codon_start=1
/transl_table=1
/product="putative tail component of prophage CP-933v"
/protein_id="AAG57206.1"
/db_xref="GI:12516370"
/translation="MATRELITIKISANSQSFQSTORASRMGSEYRTILONGGROA
AAAREORRALAELNSQLTETRGSAVAGAFAGAFASGLHLSIADENSSVARRLKQAS
OSDEFASSQKYLMDISORTGAESDNALFARSAASRMREYGSAGDYLKYETAISTG
LRTSGASTAEAKSVITTOFSQALAOGLVLEGEFFNVNSNGSRIYRALAAGVARDLK
AMADOCKLRADYVVALISQGLADEYAAAPETVSSITKYNALMAVVGANASG
VKTLSGLMNGVAGQIDNVATAVAGAAVAVGAREGMAAGASATAGLVTAAREVA
LAEOQRTQIATARAAYVRAQAAVAAARGTEKQIABARLAATOERLNRIARS
AAQNALNSTTAVGSRMKGALGLVGVPCYVMLGAAMWTLYQNOEARSARYALT
EDRRVALIRQOAAADONKYOSMVMNQHEFRNLIGLQNELLOOQGVGVYDAPLPLP
ATLDDKOOSALTTEREALSTRUKAEDEKERARLGVADDLGVGXOGEARQRTSMA
LEAMNNANKPKRGCKSETERAEQSFSLKKOXXOLATGONTPELAKIYOTAG
ELKTLETQKQELLRNALIDQOKIREOURSERETLKNEBAARSNAELGLGGE
RAREMRDQIIRDFEOKDADLOSOYOTGIDSEDPYRQALAAQNQITLSEDLKQXXF
YASDXORAPDMQGLDGFSENVWDNASDYASQAOALATGEGISGVNNTTEMLNKKYE
MRSASSVLOEISKVLMNAIVGINTAANGAGGGLGSGISGMVGAVAMKGVY
TSANLSVNSISYIDTPYPAFAKAGAGLKGAEPAIMPLTRADGSLGXRXNAV"
complement(3333..3614)
/gene="23319"
complement(3333..3614)
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 90 of 93 are 61.11 pct identical to
residues 133 to 222 of 224 from GenPept 118 :
g116901601|gb|AAJ31104.1| (AF069529) gp14 [Bacteriophage
HK97]"
/codon_start=1
/transl_table=1
/product="unknown protein encoded within prophage CP-933v"
/protein_id="AAG57207.1"
/db_xref="GI:12516371"
/translation="MRLALRLGRITLSELRLHSLSASEAMMNEFDVSLGDERQDIRN
AOIYKAVGCAQGNVNAIKDMLCWDGDEDEKPEVDFPALDELALSLAAS"
complement(3638..4012)
/gene="23320"
complement(3638..4012)
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 124 of 124 are 62.90 pct identical to
residues 1 to 123 of 224 from GenPept 118 :
g116901601|gb|AAJ31104.1| (AF069529) gp14 [Bacteriophage
HK97]"
/codon_start=1
/transl_table=1
/product="unknown protein encoded within prophage CP-933v"
/protein_id="AAG57208.1"
/db_xref="GI:12516372"
/translation="MARDIKTLALRLSGFRHXTKVPBMRVSVLREPSAEANYLM
QEVINGDEDDDTLSVAKTRRNLEADVTLLFCVLDCTDLQGVFXPDREQVLAVYGP
VHARLRQALELINDAESARRK"
complement(4018..4863)
/gene="23322"
complement(4018..4863)
/function="putative structure; Structural component (Phage
or Prophage Related)"
/note="Residues 44 to 277 of 281 are 67.52 pct identical
to residues 1 to 234 of 234 from GenPept 118 :
g116901588|gb|AAJ31091.1|AF069529.4 (AF069529) major tail
subunit [Bacteriophage HK97]"
/codon_start=1
/transl_table=1
/product="putative major tail subunit encoded within
prophage CP-933v"
/protein_id="AAG57209.1"
/db_xref="GI:12516373"
/translation="MKCRITFRKAVTVQWMSFRSCNFPNFTHNPAPGFLISGRM
SALYERSQILQWMTSSNPATETWEKAEYLRLLCTIKEQPAGGOKODIDVTTLCSTE
OENINGCASSSEISMGNFVLNOAONALRDAYNDTVAEFKYQPSGGRFKFLABVRO
HMSGNTNGVAATFSLRLKGFVSVVPLAFVKNXXKRTLVNTGALLTMSVSNCGT
NPYKXAMKKGQDPEQGTDTTFSKANKSGDKGATCYXVDSADPOSDTSDACTIV
NGAG"
complement(4793..5137)
gene

CDS	/gene="23323" complement(4793..5137) /gene="23323" /function="orf: Other or unknown (Phage or Prophage Related)" /note="Residues 1 to 113 of 114 are 49.12 pct identical to residues 1 to 114 of 115 from Genpept 118 : g1169015919b1/AAf31102.11 (AF069529) Gp11 [Bacteriophage HK97]" /codon_start=1 /transl_table=11 /product="unknown protein encoded within prophage CP-933v" /protein_id="AAG57210.1" /db_xref="GI:12516374" /translation="MTEDLPYHLAHLAAGGVYPPVPLDGRPSVALPWPVYSLISSASADWGGAESVSVDIVAGTVTAQRQITQDAKEALIMLAPGSVSBODYIPENRCYRATLEFQVY" complement(5134..5580) /gene="23325" complement(5134..5580) /gene="23325" /function="orf: Other or unknown (Phage or Prophage Related)" /note="Residues 1 to 148 of 148 are 75.16 pct identical to residues 1 to 149 of 149 from Genpept 118 : g11690159819b1/AAf31101.11 (AF069529) Gp10 [Bacteriophage HK97]" /codon_start=1 /transl_table=11 /product="unknown protein encoded within prophage CP-933v" /protein_id="AAG57211.1" /db_xref="GI:12516375" /translation="MTETLDFSGLEDISRDQLLSGAENNRVLRATAGANVLEEVSRAVRRGLKRRNVVLSRRSGDGMESGVHICGVNPTGNSDVTMTADNPRNAFYWRFVEMGTVMPPHPVPAPFVRSQQAAQVAIAIRNRAIDVLR" complement(5577..5927) /gene="23326" complement(5577..5927) /gene="23326" /function="orf: Other or unknown (Phage or Prophage Related)" /note="Residues 6 to 116 of 116 are 45.94 pct identical to residues 4 to 112 of 112 from Genpept 118 : g11690159719b1/AAf31100.11 (AF069529) putative head-tail adaptor [Bacteriophage HK97]" /codon_start=1 /transl_table=11 /product="unknown protein encoded within prophage CP-933v" /protein_id="AAG57212.1" /db_xref="GI:12516376" /translation="MAISAGRLTOMISVLPVLTNRNAGMTBEMVSGCKIHADIRGRSSRRMOSAEQAQAEIRIWRGQSGRETTAASRLHVLISGPRDXLNVVAGXPVPDXT	4.6%; Score 22; DB 1; Length 13505; Best Local Similarity 100.0%; Pred. No. 0.48; Matches 22; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;
gene	Query Match 291 TAATCCGGACACCGGTACAGC 312 Db 5338 TAATCCGACACCGGTACAGC 5317	
CDS	RESULT 10 AE006459/c 13624 bp DNA 11linear BCT 21-MAR-2001 LOCUS Escherichia coli O157:H7 EDL933 genome, contig 2 of 3, section 53 DEFINITION of 55. ACCESSION AE006459 AE005174 VERSION AE006459.1 GI:13259579 KEYWORDS SOURCE Escherichia coli O157:H7 EDL933. ORGANISM Escherichia coli O157:H7 EDL933 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	
gene	REFERENCE 1 (bases 1 to 13624) AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dinalanta,E., Potamotis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R. TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7 JOURNAL Nature 409 (6819), 529-533 (2001) MEDLINE 21074935 PUBMED 11206551 REFERENCE 2 (bases 1 to 13624) AUTHORS Plunkett,G. III. TITLE Direct Submission Submitted (12-FEB-2001) Laboratory of Genetics, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706-1577, USA JOURNAL Location/Qualifiers 1. 13624 /organism="Escherichia coli O157:H7 EDL933" /strain="EDL933" /serotype="O157:H7" /db_xref="taxon:155864" /note="enterohemorrhagic" complement(198..3440) /gene="26034" complement(198..3440) /gene="26034" /function="putative structure: Structural component (Phage or Prophage Related)" /note="Residues 1 to 1076 of 1080 are 52.37 pct identical to residues 1 to 1079 of 1080 from Genpept112: g11AF311092.11(AF069529_5 (AF069529) tail length tape measure protein [Bacteriophage HK97]" /codon_start=1 /transl_table=11 /product="putative tail component of cryptic prophage CP-933p" /protein_id="AAK16950.1" /db_xref="GI:13259580" /translation="MATLERLIKISANSQSFSQETIORASRMGSEVYRTLONCGROAA AAAREQRALAEHLISQLEIRASAVMTGAFGATGHLISLADWSSVNRILQOAS OSSDERASSOKVIMDISORTGTAESDNALFPRASAMNEYSAADVLYKEALSTG LTKISGASTAEAGSVITPFSQALACGYLREEFNSVNSGDRIVRLAACMGVARDLK AMADDKLTADYVVPALISQILRLREYAMETVSSSTTKENAPMAVGCANESAG VYKTLGSLMGVAGQIDNVATVAGVAVARYRGNMAGMSATAGLVTAARNEVA LAEAFRTQIATARAARAVYRAQQAVALAARCTEMQIAEARLATAQERLNRIARS AAONALNSTTAVGSRIMSGALGIVGPGVLMGLAAAVTLVONQBOAREASROYALT IDEIAHKTPMSLSPEASDNEGRRAALTRPONHLIDQASRVKSLOEKINGYQYVLANP GMTTGGCFEINHLTSKYTYTBELAQTEDLAYEQSRLAQMOEKAOSIODVLAGLEDR VALTROAAEONKRVYOSLMVMOQYTERNRLIGLQNELIQQRGLVNVFLRPLQATLD DKQOQSALTTERELALSRLLKGEKEKERVRLGVAADLGFVGDYQYQERORYISNALQW RNNENVKPKSRGCKSETERKEDSFSLRLOQOEOLALAQNTLELAKLYOTALGELT LSEIKOEILRNAAALIDQOKIREQLRYREETLKNVVARASNESELGYGGERARE RMRELOQIRDSFRKQADLOSOYOTGDISEDPYRLOANAOYLISRLKINDQVFAES DYORADMQKIDEGFSNWNVDNSASQAACQALATBGISMWNNTTEMLNKDNVFRSM ASSVLOEIRKVLNNAIVNGITTAAGMSAGAGFUSTIDWMLGAAVANKKGVITTSAN LSAYSNSIVDTPYFAFAFGAGIGAGEAGETIMPLTRADGSLGVRAVSNMGSAGLV VSPVYHAIQNDGANGQIGPEAAGSLVQIDQVAVVLMRMDGMLSG" complement(3488..3769) /gene="26035" complement(3488..3769) /gene="26035" /function="putative factor: DNA packaging, phage assembly (Phage or Prophage Related)" /note="Residues 1 to 92 of 93 are 60.86 pct identical to residues 133 to 224 of 224 from Genpept121: g11/AAf31104.11 (AF069529) Gp14 [Bacteriophage HK97]" /codon_start=1 /transl_table=11 /product="putative tail assembly protein of cryptic prophage CP-933p" /protein_id="AAK16951.1"	

```

/db_xref="GI:13259581"
/translation="MRLALRLGRTLSLSLSASEAMMMMEFRISPLGDDRGDIRN
AQIVKAVGAGGMNVALKDAMLGWGEDEKPEVDPPALEDLNLSLAQS"
gene
CDS
    complement(3793..4215)
    /gene="26036"
    complement(3793..4215)
    /function="putative factor: DNA packaging, phage assembly
    (phage or prophage related)"
    /note="Residues 17 to 140 of 140 are 62.90 pct identical
    to residues 1 to 123 of 224 from Genpept121:
    gb|AAE1104.1| (AF069529) Gp14 [Bacteriophage HK97]"
    /codon_start=1
    /transl_table=11
    /product="putative tail assembly protein of cryptic
    prophage CP-933p"
    /protein_id="AAK16952.1"
    /db_xref="GI:13259582"
    /translation="MHRDSERSRPRRGVINKKDKTLALARLSGFRHKTVKPEMGN
    VSVLRPEPSAEAWYLMQVINGDEDDDTLSVAKTRRNLEADVTLFCVYLCDDTLDR
    VTPDDREOVAVYGPVHARLLKQALSLIADAESARK"
    gene="26037"
    complement(4182..4967)
    /gene="26037"
    /function="putative structure: Structural component (Phage
    or Prophage Related)"
    /note="Residues 24 to 257 of 261 are 68.37 pct identical
    to residues 1 to 234 of 234 from Genpept121:
    gb|AAE1091.1|AF069529_4 (AF069529) major tail subunit
    [Bacteriophage HK97]"
    /codon_start=1
    /transl_table=11
    /product="putative tail component of cryptic prophage
    CP-933p"
    /protein_id="AAK16953.1"
    /db_xref="GI:13259583"
    /translation="MPSFPRYKTHPPRAGFLSLGGRMSALYERSOULTQVMSISAT
    LNOAALRDADNDALIAFVLEPSSGGEFLAEVROHWTSSGNGVAAAFPSLTK
    GKPVSFVPLAFVKNLDTLTNTGALLTMSVSAAGCTPPYKRAMKKDDQPDVGQRTK
    TESKGAOSADGKTYCTVVDISAERKASVSVECTYVSAAG"
    gene="26038"
    complement(4964..5308)
    /gene="26038"
    /function="putative structure: Structural component (Phage
    or Prophage Related)"
    /note="Residues 1 to 113 of 114 are 49.12 pct identical to
    (AF069529) Gp11 [Bacteriophage HK97]"
    /codon_start=1
    /transl_table=11
    /product="putative structural component of cryptic
    prophage CP-933p"
    /protein_id="AAK16954.1"
    /db_xref="GI:13259584"
    /translation="MTEADLPPLAHLAGGVYVYVLLDGRPSVALPWVYSLISS
    VSADVMGGAESSVSQIDVYAGVTQARQIRQARAIMLAPGVSEMODYIPENR
    CYRATIEFQVY"
    gene="26039"
    complement(5305..5751)
    /gene="26039"
    complement(5305..5751)
    /function="putative structure: Structural component (Phage
    or Prophage Related)"
    /note="Residues 1 to 148 of 148 are 75.16 pct identical to
    (AF069529) Gp10 [Bacteriophage HK97]"
    /codon_start=1
    /transl_table=11
    /product="putative structural component of cryptic
    prophage CP-933p"

```

```

/gene="26040"
/translation="MIEHLDFSGLEIDISRDLLLSGAEENRVLREATPAGANVLEE
WFEVMEGVNMPPHPFVRPAPDVASDAQVQALARMNAIDEVLR"
complement(5748..6005)
/gene="26040"
complement(5748..6005)
/gene="26040"
complement(5748..6005)
/function="putative structure: Structural component (Phage
or Prophage Related)"
/note="Residues 6 to 85 of 85 are 53.75 pct identical to
(AF069529) putative head-tail adaptor [Bacteriophage
HK97]"
/codon_start=1
/transl_table=11
/product="putative head-tail adaptor of cryptic prophage
CP-933p"
/protein_id="AAK16956.1"
/db_xref="GI:13259586"
/translation="MSCGKIHADIRGSSRRRMOGAEMAOEIRIWRGSGRETTA
ASRLHVLSPWMDRLINVGLPVPDATTGRLTLCRLGKER"
complement(6108..6434)
/gene="26041"
complement(6108..6434)
/gene="26041"
complement(6108..6434)
/function="orf: Other or unknown (Phage or Prophage
Related)"
/note="Residues 5 to 25 of 108 are 66.66 pct identical to
gb|AAE27360.1|AF198256_14 (AF198256) unknown [Haemophilus
influenzae]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933p"
/protein_id="AAK16957.1"
/db_xref="GI:13259587"
/translation="MTALITLFEIKRAHLDVNDADDDMLNDKVRQATAVLLAYIGSSR
DKVIRDEGELIPEALTRMKAGAMRLTGMVLRNDLAERBELIQGELPSSVSLIIDL
RCPTVY"
complement(6437..9016)
/gene="26042"
complement(6437..9016)

```

Query Match	Best Local Similarity	Score	DB	Length
Matches 22: Conservative	100.0%	0	48	0
Matches 22: Mismatches	0	0	Indels	0
Matches 22: Gaps	0	0	0	0

```

QY 291 TAATCCGACACCGGTACACG 312
Db 5509 TAATCCGACACCGGTACACG 5488

```

RESULT 11
 AL139161
 LOCUS
 DEFINITION Human DNA sequence from clone RPS-940F7 on chromosome 1q42.2-4.3,
 complete sequence.
 ACCESSION AL139161
 VERSION AL139161.8
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 103608)
 AUTHORS Martin, S.
 TITLE Direct Submission
 JOURNAL Submitted (04-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk; Clone
 requests: clonerequest@sanger.ac.uk

COMMENT

On Aug 9, 2001 this sequence version replaced g1:14787245. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL, Sw: SWISSPROT, Tr: TREMBL, Wp: WORMPEP; information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> RP5-940F7 is from the library RPCI-5 constructed by the group of Pleter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP5-940F7. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP4-764D2 is at 101609 in this sequence. The true right end of clone RP4-580N22 is at 2000 in this sequence.

FEATURES

source

1. 103608
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q42.2-q3"
/clone="RP5-940F7"
/clone_lib="RPCI-5"
1. 151
/note="AluSg/X repeat: matches 145. .295 of consensus"
369. .935
/note="L1PB2 repeat: matches 5574. .6155 of consensus"
936. .1247
/note="AluSp repeat: matches 1. .313 of consensus"
1248. .3222
/note="L1PB2 repeat: matches 3614. .5574 of consensus"
2001. .2044
/note="Single clone region. Assembly confirmed by restriction digest data."
3462. .3798
/note="MER58B repeat: matches 2. .341 of consensus"
4257. .4555
/note="AluSc repeat: matches 5. .298 of consensus"
4574. .4663
/note="AluSc repeat: matches 161. .248 of consensus"
4714. .4753
/note="L2 repeat: matches 161. .248 of consensus"
4812. .4870
/note="MER4D repeat: matches 359. .417 of consensus"
4870. .5107
/note="MER4D repeat: matches 9. .234 of consensus"
5895. .6160
/note="AluX repeat: matches 6. .285 of consensus"
6161. .6196
/note="L2 repeat: matches 2 mer ca 97% conserved"
6290. .6402
/note="L2 repeat: matches 2588. .2705 of consensus"
6837. .6868
/note="L2 repeat: matches 2 mer ca 84% conserved"
7155. .7456
/note="AluX repeat: matches 1. .303 of consensus"

repeat_region 7653. .7833
/note="MIR repeat: matches 33. .208 of consensus"
8397. .8646
/note="L2 repeat: matches 2 mer tg 100% conserved"
8954. .9255
/note="AluJ repeat: matches 1. .301 of consensus"
9373. .9506
/note="L2 repeat: matches 2606. .2733 of consensus"
10071. .10761
/note="L1MC4 repeat: matches 7162. .7974 of consensus"
10847. .11024
/note="MIR repeat: matches 59. .229 of consensus"
11025. .11251
/note="AluJ repeat: matches 74. .304 of consensus"
11275. .11414
/note="AluJ/FRM repeat: matches 147. .286 of consensus"
12228. .12536
/note="AluY repeat: matches 1. .309 of consensus"
12537. .12829
/note="AluJ repeat: matches 11. .300 of consensus"
13542. .13637
/note="MIR repeat: matches 46. .143 of consensus"
14166. .14457
/note="AluJ repeat: matches 1. .292 of consensus"
15312. .15622
/note="AluX repeat: matches 1. .306 of consensus"
16455. .16515
/note="L2 repeat: matches 2647. .2708 of consensus"
17415. .17458
/note="L2 repeat: matches 2 mer ca 79% conserved"
17480. .17949
/note="MIR repeat: matches 1. .466 of consensus"
17950. .18049
/note="L2 repeat: matches 1. .466 of consensus"
18444. .18636
/note="MER5A repeat: matches 1. .189 of consensus"
18759. .18895
/note="MIR repeat: matches 53. .193 of consensus"
19595. .19610
/note="L2 repeat: matches 2642. .2653 of consensus"
19611. .19921
/note="AluSg repeat: matches 1. .308 of consensus"
19922. .19993
/note="L2 repeat: matches 2653. .2750 of consensus"
20033. .20154
/note="L2 repeat: matches 2 mer ca 80% conserved"
21253. .21557
/note="AluJ repeat: matches 1. .302 of consensus"
21740. .21793
/note="L2 repeat: matches 2638. .2688 of consensus"
21799. .22106
/note="AluSg repeat: matches 1. .308 of consensus"
22112. .22408
/note="AluJ repeat: matches 1. .295 of consensus"
23607. .23906
/note="AluSp repeat: matches 1. .305 of consensus"
24101. .24403
/note="AluJ repeat: matches 1. .294 of consensus"
25145. .25560
/note="L2 repeat: matches 1. .388 of consensus"
25713. .26011
/note="AluSp repeat: matches 1. .299 of consensus"
26525. .26570
/note="L2 repeat: matches 2 mer tg 76% conserved"
26573. .26859
/note="AluY repeat: matches 1. .287 of consensus"
26862. .27150
/note="AluX repeat: matches 1. .290 of consensus"
27151. .28279
/note="L1MC2 repeat: matches 5199. .6329 of consensus"
28500. .28621
/note="L2 repeat: matches 2580. .2706 of consensus"
28659. .28872

```

repeat_region /note="AluIo repeat: matches 85. .287 of consensus"
29508. .29715 /note="MER30 repeat: matches 1. .230 of consensus"
repeat_region 29785. .30063 /note="AluIo repeat: matches 1. .250 of consensus"
repeat_region 30640. .30754 /note="LIM4 repeat: matches 4494. .4608 of consensus"
repeat_region 30762. .30933 /note="MLTIC repeat: matches 1. .182 of consensus"
repeat_region 30942. .31167 /note="LI repeat: matches 4493. .4721 of consensus"
repeat_region 32192. .32894 /note="MER50 repeat: matches 1. .711 of consensus"
repeat_region 33194. .33485 /note="AluSx repeat: matches 1. .301 of consensus"
repeat_region 33563. .33697 /note="AluDb repeat: matches 1. .136 of consensus"
repeat_region 33698. .33995 /note="AluSg1 repeat: matches 1. .297 of consensus"
repeat_region 33996. .34174 /note="AluDb repeat: matches 136. .312 of consensus"
repeat_region 34670. .34769 repeat: matches 2361. .2461 of consensus"
repeat_region 34927. .35112 /note="LIR repeat: matches 48. .243 of consensus"
repeat_region 35604. .35905 /note="AluSc repeat: matches 1. .302 of consensus"
repeat_region 35906. .36181 /note="AluSg repeat: matches 1. .277 of consensus"
repeat_region 36585. .36625 /note="MER46 repeat: matches 196. .236 of consensus"
repeat_region 36626. .36797 /note="LIR repeat: matches 1518. .1346 of consensus"
repeat_region 36795. .38360 /note="LIR repeat: matches 184. .184 of consensus"
repeat_region 38361. .38657 /note="AluSg repeat: matches 1. .302 of consensus"
repeat_region 38658. .41845

```

```

Query Match
Best Local Similarity 100.0%: Score 22; DB 9; Length 103608;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 257 AAAAAGTGAAGCGCGCAGCT 278
Db 7141 AAAAAGTGAAGCGCGCAGCT 7162

```

```

RESULT 12
AP002555 222605 bp DNA linear BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 6/20.
ACCESSION AP002555 BA000007
VERSION AP002555.1 GI:13360886
KEYWORDS
SOURCE
Escherichia coli O157:H7 (strain:O157:H7, sub-strain:RIMD 050952)

```

```

ORGANISM
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (sites)
Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Yamamoto,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Sasakawa,C., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Complete nucleotide sequence of the prophage vW2-Sakai carrying the
derived from the sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)

```

```

JOURNAL MEDLINE 20198780
REFERENCE 2 (sites)
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and

```

```

TITLE
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
JOURNAL MEDLINE 20557356
REFERENCE 3 (sites)
Yokoyama,K., Makino,K., Kubota,Y., Matanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C.,
Shinagawa,H.
Complete nucleotide sequence of the prophage vW2-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
4 (sites)
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsuno,E., Nakayama,K., Murata,T.,
Tanaka,M., Toke,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 222605)
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:ken-gen-info.osaka-u.ac.jp, Tel:81-6-6879-8365,
URL:http://www.gen-info.osaka-u.ac.jp, Fax:81-6-6879-2047)
genome project.

```

```

COMMENT
FEATURES
SOURCE

```

```

gene
CDS

```

```

gene
CDS

```

```

/location/Qualifiers
1. .222605
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 050952"
/db_xref="taxon:83334"
74. .607
/gene="ECS1423"
74. .607
/gene="ECS1423"
/note="similar to B1045.ECOLI g11787283 percent identity
100 in 177 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative polyprotein"
/db_xref="GI:13360886.1"
/translation="MKTRIHVQGDITKLAVDYVYNNANSLMGSGGVDGAIHRAAP
NSRLVYANASYSVAIPAISTGVGYPRAAAEIAVATSEFTIRHALPEQYFYCYD
EENALYERLILQCGDE"
549. .2030
/gene="ECS1424"
549. .2030
/gene="ECS1424"
/note="similar to YMDC.ECOLI g11787284 percent identity
99 in 493 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative synthase"
/protein_id="BA334847.1"
/db_xref="GI:13360888"
/translation="MKKTPSTKDSLPNKEENDLPRLASVYPLCSQHPGCGGLPEPL
EKSLDPAARYRYRAEMSEHTLDVQYIWMDDMSGLIFSLALAAKAGVAVRLLDGN
NTPGCDLIRLDSHRIEVRLEPNPFRLLRPLIGYITDPSRLNRHMKNSFTVDGVN

```

```

TLVGRNIGDAYGAGEEPLFSDLDVMAIGVVEDADDFARYMCKSVSPLOQVLDV
PEGEMADRIELPASHMNDAMTHRYLKRMESSPEINHLVDGTLPLIMAKRRLSDDAK
GGGAKRHSLLPORLEFDIMGSPSERIDITISYFVPRACVAGOLIMVRRGVKATILTN
SLAANDVAVYHAGYARMRKKLLRGVELYELKPTROSSLTHDRGTGSGSLHAKT
FSDIGKTVFTIGSFNDFPNSTLNTMGVFIESTLQOLIDKRFIDISOYDAAMOLRIDR
WGRINMVDNRHAKKEIYLKRETPATSFMRKRVMLASILPVEMLL"
/gene="ECS1425"
/complement(2038. .3195)
/gene="ECS1425"
/complement(2038. .3195)
/gene="ECS1425"
/notes="similar to MDOC_ECOLI g113025226sp1P75920 percent
identity 99 in 385 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="glucans biosynthesis protein"
/protein_id="BAB34848.1"
/db_xref="GI:13360889"
/translation="MNPVPAOREEYFLDSIRAWMLLGIPEHISLISYSHTHVNSAP
SLMLTLFNDHFHSFRMQVPEFISGVSYMLPLRYPLKMKMKVREVRGIDMLTAIPLL
TLPOPIMLOYVKAESMPGLSIDYKNTLANELISHMLFLVLYVMTLCAWIFERI
RNNLENSDKTRKFSMWKLSVIFLCLGIGAVIRITIFIVPEILSNGFNTVMQTL
EYLPEFIIICALAFIEPHILALFTTPSGCTLAALAVAYLNLQRYGSDAMWYETES
VITMWGLMMWVVSFGRHLNFOGARVTVNANSLFTYLVHHPLETLFEFGAYTPPHI
TSNMLGFLGILFEVVGIAIILVEIHRIPLKFLFGSKPVKRENDKABAR"
/3588. .5123
/gene="ECS1426"
/3588. .5123
/gene="ECS1426"
/notes="similar to MDOC_ECOLI g11787286 percent identity
99 in 511 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="periplasmic glucans biosynthesis protein"
/protein_id="BAB34849.1"
/db_xref="GI:13360890"
/translation="MKMKRMLSAAMVLTLYTSSNAFSIDVAKOKOSLAGKEAPK
SNLPSYFRMKYADVOOIQPNHDKAMNMLKPEFIEFHOCGRPTPVKINVTYTA
VKRIKISPYFTFGDVQHDKDTYKDLGFGFVKLYTINSKDKDELYVNLGASIFRYI
KAGQVYGLSARGLAIDTALPSEGEPRFEFWLERKPTDKRLTYIALDSEPRATGAY
KEVVMGCRDTVVDOVSKILYLRDKVGRKLGVALPSMLFGPNQSPANNYRPELHDSG
LSIHAGNGEMIRPLNNPKHLVASSFSMPOGFGLLQGRDSEFREDLDDRYLDRPS
AAVTPKGEWKGKSVLELVEIPTNDETMNIVAAWTPPOLDPGCKEMNKXTITFSRED
KLHAPDNAAVOCTRSTGDYKOSNLIRQDGTIAPVVDYTGAKMKLPPDTPYATATS
IGNGETIVESTVRYNPTVTKMRLVAKVKADAKTTEMPALVADQTLSETWSIOLP
ANE"
/5086. .7659
/gene="ECS1427"
/5086. .7659
/gene="ECS1427"
/notes="similar to membrane glycosyltransferase MDOH_ECOLI
g11787287 percent identity 100 in 847 aa but differs at
N-ter (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="membrane glycosyltransferase"
/protein_id="BAB34850.1"
/db_xref="GI:13360891"
/translation="MKPGATSYLPNNKTTTEYIDAMPJAASEKALPKPTDIRAHOALD
AHRPMARREDSDPOGSKARLEQAMPDSLADQDLIKDDGGRQOLKMPAKKSSMPED
PKRTNPGFRMRLRGDRDYTPRIALRTREBDESBDKMTYGTGIRYITILLITLQATV
VATWIKTILPLTQGNALINPMQVGGDLVWSEQLPLMLQGTILITLFLVLEWVSAG
FWTALMGFLQDLIGRDKYSISASTVGDDEPLNPHRTALIMPICNEVNFVGLRWTA
ESYKATGNAKKHDFYILSDSYNDICVAOAKAMETLAEVGGQGFYRRRRRVK
SGNIDFCRRMSQSYVYVLDASVTGDCGLGRLMEANPNAGIIOSSPRASSD
TLYARCOQCATRYTGYPLTAGLHFWQLGSHWGHAAIIRVPEPIHCLALPAGGGS
FAGSILSHDFVEALMRACMKWITAYDPCSEDELPPULDELKDRKRWCHNLMNF
RLFLVKGHPHRAVFLTGVMSYLSAPLWFMFLAUSTAIAQVHALTEYVLTSLLEVES
PWPQWPRELAIALFASTVWLLFLPKLLSLILWCKGTEYGGFWMTVLTSLLEVES
VLLAPVRLMFLHYFVVSARFLGWVWNSPQRDDSTSMGEARKRHSQLLGLVMAVG
MMLDURFLFLWLAIVFSLILSPVSVISSRAIVGRTKRMKFLITPEVSPPOVLVD

```

```

TDRLFLEMRQSLDDGFMAHAFNPSEFNALATAMATRRHRSKYLEIARDRIHEQALNE
TPEKLRDRRLVLLSPVPMARLHFHFWNWSPEYSSWVSYGIKLNPALAKKPDAA5
0"
/gene="ECS1428"
/7832. .8059
/gene="ECS1428"
/7832. .8059
/gene="ECS1428"
/notes="identical to YCEK_ECOLI g11787288 (Conserved in
E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34851.1"
/db_xref="GI:13360892"
/translation="MKLIVSVMWTLISCCGSIISRTIGCGHGNQYRGVQVDYRDS
ARKRYVITDLPESLVFDTLLPLDIDHNGYE"
/complement(8060. .8437)
/gene="ECS1429"
/complement(8060. .8437)
/gene="ECS1429"
/8437)
/gene="ECS1429"
/notes="similar to MSYB_ECOLI g11787289 percent identity
99 in 125 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="acidic proteinin MsyB"
/protein_id="BAB34852.1"
/db_xref="GI:13360893"
/translation="MWTMYATLEAIDAAREFLADNPQIDAEADANVOQFNQKYYLQ
DDDIMVQVEFPADEGECEGLPMLSGEAAQSVFDGDDYDEIIRQENQENTJLHMEDEG

```

Query Match 4.68; Score 22; DB 1; Length 222605;
Best Local Similarity 100.0%; Pred. No. 0.55;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 291 TAAATCGGACACCGGTAACAGC 312
Db 105175 TAAATCGGACACCGGTAACAGC 105196
|||||
|||||

RESULT 13
AP002557/c 286485 bp DNA linear BCF 07-MAR-2001
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE

AP002557 286485 bp DNA linear BCF 07-MAR-2001
Escherichia coli O157:H7 DNA, complete genome, section 8/20.
AP002557 Ba000007
AP002557.1 GI:13361466
Escherichia coli O157:H7 (strain:0157:H7, sub_strain:RIMD 050952)
DNA.
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (sites)
Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tetsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C., and Shingawa,H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
2 (sites)
Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shingawa,H., and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
3 (sites)

AUTHORS	Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Shinagawa, H., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Complete nucleotide sequence of the prophage VT1-Sakai carrying t
JOURNAL	0157:H7 strain derived from the Sakai outbreak
MEDLINE	Gene 258 (1-2), 127-139 (2000)
REFERENCE	4 (sites)
AUTHORS	Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, S., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.
TITLE	Complete genome sequence of enterohemorrhagic Escherichia coli
JOURNAL	0157:H7 and genomic comparison with a laboratory strain K-12
MEDLINE	DNA Res. 8 (1), 11-22 (2001)
REFERENCE	21156231
AUTHORS	5 (bases 1 to 286485)
	Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.
TITLE	Direct Submission
JOURNAL	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
	Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
	Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL: http://www.gen-info.osaka-u.ac.jp,
COMMENT	Fax: 81-6-6879-2047)
FEATURES	genome project.
SOURCE	1 location/Qualifiers
	1 286485

```

98  in 879 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB35426.1"
/db_xref="GI:13361469"
/translation="MICKKRVALLALLLLEPLLMLTGLMVPPTLAGTGLPLG
ALDESPTTRKGLIPDLRVLVSDCCILAIITNASSHSRNLNGLVLEIDSKCL
PTEGSPAPKPTKMSQMSMLPNTWINDILLESPOMGOSRLNLSLALSDIQQLRYG
KKEFGGKLGSGGLTVSELDVLANEENSPKVLKEFTMLMLDGLIPSGHATVLTGL
ESLDELDELQENSGGLVLVLANENGDPLTLPQGITPQGLTVSDGMSVPAFGF
GRVCAVQNMQAGLENALVSGRLVLTGQAGKGNAYLNGPGLSDMSNSQLPTLQ
EAKQDILILRLPLDGLSGLSPLTTEPGALLRSGRVDSLDJDEIRMLPAJLGL
TORVDSRDLAIIQAENELGDFVLQMGDLNGLPAGRMQKRVYKSGSEFTPMANA
DYAKGSGNLTITLIDLSLGEFLOLQTMVETPKLLDKPVVNRDVGNDGTR
SLDGQSLTFLGSGVLPEPTLTKPSVNGRPPTFLKGLDLHGLGIVRVNGRMDGR
GNMWPVKSLTVAPQLVPPDMKMLRODELTAQVAFSAEAGFEFGAGGVLTKGSS
MPNQNQVNDVLPVPPFADGAMHIGTRPVTLLRAEVLINLTAKNTVDLIGCRVY
SEBELLTLTVSEDLGNTGLMKLQEPHLDALRLNNLSSELVASVNDKOPANSG
SGALLPILNKKKYLTPQANGLANAGPNTLRKTDJADAVAKDMWTGTAIIINLRV
RSTLRINDINLCVLTQKQNLNAGPNTLRKTDJADAVAKDMWTGTAIIINLRV
WLEQNAQLPETHCGSGKCEERKQ"
/length=4845
/genbank="gbss2004"
/seqid="852004"

```

```

//evidence=not_experimental
//product="hypothetical protein"
//protein_id="BAB35427.1"
//db_xref="GI:1361470"
//translation="MKIVAAPASSFMLVGTPTIEVAAPKEPITINNKKVLEHII
KAKVDIEDILKRSDF"
4830..5176
//gene="ECS2005"
DSDS

```

gene	5590..8625	5590..8625
CDS	/gene="EGS2006"	5590 8625

gene
CDS
"
/product="fermentative D-lactate dehydrogenase"
/protein_id="BAB35425.1"
/db_xref="GI:13361468"
/translation="MKLIAYSTKQDKKRYLQOVNESFGPELFFDFPFLILETKTANG
CEAVCIFVNDGDSRPYLEBKLRKHGVYIALKCAGFNVNDIDAKELGLKVRVAYDP
RLAEVHEHIGMMATLNRIIRAVQRTRODNESLGIDGFMGYKTHGIVGTGKIOVAL
AFLFGPMRLAVDPYSAALGLGEVEVDPLTESQVYISLCHKPLTPENYHLILNA
ODDYFRLSLCHNIVFLTQHQAFLEALVISTITLQNLINLKEGTCPELV"
2024..4663
/gene="ECS2003"
2024..4663
/gene="ECS2003"
/note="similar to YDBH_ECOLI gi|11787466 percent identity

```

/gene="ECS2005"
/note="similar to YDL_ECOLI g11787648 percent identity
71 in 109 aa but differs at both N-ter and C-ter"
/codon_start=1
/transl_table=11
/evidence=no_c_eperimental
/product="hypothetical protein"
/protein_id="BAB35428.1"
/db_xref="GI:13361471"
/translation="MMKRTLLIVNAIGMDSAMALTLEARSOGREVFETNGTLYVA
LQTDLEFALVKDINERNHSTOOLAKONNVSTKEIRAGQLVARKSQDYQGVN
GWLNR"
5590..8625
/gene="ECS2006"
5590..8625
/gene="ECS2006"
/note="ECS2006"
/note="unknown, similar to Biga protein [Salmonella typhi]
dad|AF131385.1|AAD9677.1 percent identity 32 in 1006 aa,
similar to N-terminal part (1-1128 in 2020 aa) of ydaA2'
[Escherichia coli] glf17465766|ptr||C48399 percent identity
49 in 1011 aa"
/codon_start=1
/transl_table=11
/evidence=no_c_experimental
/product="putative biga-like protein"
/protein_id="BAB35429.1"
/db_xref="GI:13361472"
/translation="MECKKKLSYCAWMAISQTAAADTSIPDSTEKRKSITCPAN
HISLKEDLENLSAQECRKLSDYPAALIAAVATGAIITYLSDDDHHNNSSPVP
DDGGTPVPDDGGTPVPDDGGDTVPDPDDGGDTVPDPDDGGDTVPDPDDGGTVP
DDGGTPVPDHNVKNVYKNSVTWDQAQKVLRFTTFEYSNDGSGTYTLAEFGETLV
KSMVYHDNSNMVDEGNTVSGSITWSYDDGLIHITFEAGVVVDSTGTALAEFNAL
TDCGNALMGSLVMTWDDGDNISLNDDGKTTAIGESVVCILTGSNITINNGETED
```


transport proteins e.g. g1|3661478|9b|AAC61710.1| percent
 identity 91 in 284 aa
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="molybdenum transport protein"
 /protein_id="BAB35117.1"
 /db_xref="GI:13361159"
 /translation="MIFLSQADIALLEDJQSGDLTTALNIGHQGYIEPIRQGG
 CQGVSVACKMLTTGLTIDAVSPGSOANAGORLIRAGCNAAALHQKKAIONLEW
 SCGVSDYLDOMALILRRYPGNAGTRKAIPTGRILASQALIAAGLTHRGACETI
 LLEFANRRHLHNDOMSCAIIQOLSRHAPKRIYVEADTPKEAIIARQPDVLDKRF
 SPQATEIAQAPSLAPPCITLALGGINLTLLKNVLDCCGIRLFTISAPYAAPADIKV
 SIQPAASY"
 complement(3219..3500)
 /gene="ECS1695"
 complement(3219..3500)
 /gene="ECS1695"
 /note="unknown"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="hypothetical protein"
 /protein_id="BAB35118.1"
 /db_xref="GI:13361160"
 /translation="MPTILSYDKTANRITAPRVLSRTSAGVLAIRLTMSVSGYIAG
 INNMKLVPSPPLPAATGRSSGIAVRRRCDFPSSGYNRCVTSGRYPGR"
 complement(3445..4089)
 /gene="ECS1696"
 complement(3445..4089)
 /gene="ECS1696"
 /note="unknown, similar to hypothetical proteins e.g. Orf2
 |Escherichia coli CPT073| g1|3661479|9b|AAC61711.1|
 percent identity 98 in 214 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="hypothetical protein"
 /protein_id="BAB35119.1"
 /db_xref="GI:13361161"
 /translation="MTIDIDIRADLYLOOLKIAHRTKTPDHMPORAEKMAENCASPT
 DSYLOOLIAKIDLCQKQTLFDMGCGPGVSLADKLTTTYGNDYSGGLNVARAA
 ALKADNVHLIORMMEDMSDLPRCDIAVASRSTLVADROMASRLNNOALRYTTHL
 VSTRFVSPAIORAKREVIETLPNYIFALNLYONGITAHYDFTRGNCODNST"
 complement(4099..4857)
 /gene="ECS1697"
 complement(4099..4857)
 /gene="ECS1697"
 /note="probable ferric enterobactin transport, similar to
 ferric enterobactin transport ATP-binding protein
 |Escherichia coli CPT073| g1|3661480|9b|AAC61712.1|
 percent identity 79 in 148 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="putative ferric enterobactin transport
 ATP-binding protein"
 /protein_id="BAB35120.1"
 /db_xref="GI:13361162"
 /translation="MTLLAVRHASLGYSRHPVLDVSFTLSOGTICLLGANGCQYTT
 LMSISGIVPILKGEVLDSLPVOTLSHROAQAIAVPOAHGCIITFESVLDVWLMGL
 APTIGAFVPGKQERKATQELKIGITLLAARWNTLSGEGRLALIALAVQDRL
 LILIDPASSLDLGHQIOLIDLTAOLKNGKMTLMSTHPLHANAIDSIIIOVEPDGRV
 TQCLPTEDLTITKLAALRYVSADQIHHLHLSAISH"
 complement(4854..5834)
 /gene="ECS1698"
 complement(4854..5834)
 /gene="ECS1698"
 /note="probable ABC transporter protein (permease), similar
 to ABC transporter permeases e.g. |Haemophilus influenzae|
 aa|2501391|sp|Q57130|YE71_HAEIN percent identity 40 in 323
 /codon_start=1

Query Match 4.68; Score 22; DB 1; Length 291136;
 Best Local Similarity 100.0%; Pred. No. 0.56;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 QY 291 TAATCCGACACCGGTACACG 312
 DB 101095 TAATCCGACACCGGTACACG 101116

RESULT 15
 AGHGBEHA/c
 LOCUS
 DEFINITION
 Agoseris heterophylla partial 18S ribosomal RNA (18S rRNA) and 25S
 ribosomal RNA (25S rRNA) genes; complete 5.8S ribosomal RNA (5.8S
 rRNA) gene; complete internal transcribed spacers 1 and 2.
 VERSION
 L13954
 KEYWORDS
 18S ribosomal RNA; 25S ribosomal RNA; 5.8S ribosomal RNA.
 SOURCE
 Agoseris heterophylla DNA.
 ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
 Agoseris.
 1 (bases 1 to 702)
 Kim, K.-J. and Jansen, R.K.
 TITLE
 Comparison of phylogenetic hypotheses among different data sets in
 dwarf dandelions (Krigia): additional information from internal
 transcribed spacer sequences of nuclear ribosomal DNA
 JOURNAL
 Plant Syst. Evol. (1993) In press
 FEATURES
 source
 location/Qualifiers
 1..702
 /organism="Agoseris heterophylla"
 /db_xref="taxon:19017"
 <1..32
 /gene="18S rRNA"
 /product="18S ribosomal RNA"
 1..32
 /gene="18S rRNA"
 33..278
 /product="internal transcribed spacer 1 (ITS1)"
 279..442
 /gene="5.8S rRNA"
 /product="5.8S ribosomal RNA"
 279..442
 /gene="5.8S rRNA"
 443..664
 /product="internal transcribed spacer 2 (ITS2)"
 665..2702
 /gene="25S rRNA"
 /product="25S ribosomal RNA"
 665..702
 /gene="25S rRNA"
 151 a 187 c 174 t 1 others

BASE COUNT
 ORIGIN

Query Match 4.3%; Score 21; DB 8; Length 702;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 GCGTGCCCTCAGCCGATGGC 456
|||||
Db 425 GCGTGCCCTCAGCCGATGGC 405

Search completed: October 27, 2002, 21:52:56
Job time : 1195.58 secs

THIS PAGE BLANK (USPTO)

RESULT 2

ID	PRELIMINARY:	PRT:	148 AA.
OC	054381;		
DT	01-JUN-1998 (TREMBLrel. 06, Created)		
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)		
DT	01-NOV-1998 (TREMBLrel. 08, Last annotation update)		
DE	17 KDA COMMON-ANTIGEN (FRAGMENT).		
OS	Rickettsia sp.		
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;		
OC	Rickettsiaceae; Rickettsiinae; Rickettsia.		
OC	NCBI_taxid=789;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=98087556; PubMed=9425244;		
RA	DAVIS M.J., Ying Z, Brummer B.R., Pantoja A., Ferreira F.H.;		
RL	"Rickettsial relative associated with papaya bunchy top disease.",		
FT	Microbiol. 36:80-84(1998).		
DR	EMBL: U76907; AAC02809.1; .		
FT	NON_TER	1	
FT	NON_TER	148	
FT	SEQUENCE	148 AA.	155FO.UT.

Query match

Best Local Similarity	36.58;	Score	303.5;	DB	2;	Length	148;
Matches	55;	Conservative	40.1%;	Pred.	No.	9.2e-20;	

25 SRQ1

25 SRQEVGATGAVGVAGOLFEGKSGPYVMATGQV
9; Caps 2

09 141 EYGTACPOPDGRWVT 157
DB 77 SORALEAPSGSSVQWRNPDNCGNYGTVPISKAY-----KNNTGQYCREYQTQTVVWGKQ 131

Db 132 K

D6 132 KAYGTACRQPDGQWV 748

RESULT 3

RESULT 3

DT	01-MAR-2003	
AC	Q9F9F2;	PRT; 159 AA
ID	Q9F9F2	PRELIMINARY;

01-DEC-2001 (Tremblrel. 19, Last sequence update)
17 KDA GENUS-COMMON ANTIGEN. Last annotation update)
Rickettsia felis (Tremblrel. 19, Last sequence update)

Rickettsiales; Rickettsia
Rickettsiaceae; Rickettsia
NCBI_TaxID=42862;

MOQUINCEFROM N.A.
MEDLINE-21217364; PubMed-11321078;
Bouyer D.H., Stenos J., Crocquet-Valdes P., Moron C.G., Popov V.L.,
Zavala-Velazquez J.E., Roll L.D., Stohard D.R., Azad A.F.,
Walker D.H.,
"Rickettsia felis: molecular characterization of a new member of the
spotted fever group.",
Int. J. Syst. Evol. Microbiol. 51:339-347(2001).
EMBL: AF195118; AM28452.1; -
SEQUENCE

Query Match

Best Local Similarity	34.08;	Score 283;	DB 2;	Length
Matches	61;	Pred. No. 6.9e-18.		

OY 8 SLLIIIV---FLVCG--AQNFSROEVGAATAVVGVAQLPEKSGSRVMAIG-CAVL 61
Db : : : : :
5 SKTITALLASMLQACNCPGPKMKKGOTGLTGAGAGALLSQCFCKKGCQL-VGVGVALL 63

Qy	62	GGLIGSGTIGOSMDOOK-----KLNOSLEKVKAGOVTRWNPDPGTSYSVEEVRTOYRN	117
Db	64	GAIVGSGIGGMDDEDRIAEIETSRALAEITPSTSEWRNPDGNSHGVTFNKTY-----	119
Qy	118	KQERROOYKREKQKAMJLAGOSKOEIYGTACPDQDGMOWIS	158
Db	120	-RNSTQVCYREYTOYVVGKQKRAKGMACRQPDGLMIOYVN	159

RESULT 4

AC		
OS2532	PREDIMINARY;	PRT; 137 AA.
01-JUN-1998	(TEMBLrel, 06, Created)	
DT	01-JUN-1998	(TEMBLrel, 06, Last sequence update)
DT	01-DEC-2001	(TEMBLrel, 19, Last annotation update)
DE	17 KDA ANTIGEN (FRAGMENT).	
Bacteriostella cooley.		
OOC	Bacteriostella proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsia	
NDBI-TaxID=69410;		
[1]		

Billir

Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Query Match	Score	DB 2	Length
Best Local Similarity	31.18;		
Matches	39.38;		
53: Conservation	Pred. No. 9e-16;		

	Indels	Gaps
22 QNFSRQEVGATGAVVGGVAGOTFCVGGG...	11;	4;

```

      /  RGMNKGSTGTLTGAGALLGSQFGKGRQL-VGVGVGALLGAVLGGQIGACMBQDRRL  78
79 --TKINOSTPRTVTRRQV--

```

66 AELTSORALEAPSGSSTEWNRPNMGNCVCYDEPWW : ||||| :
 : : | | | | | : : : | :
 : : | | | | | : : : | :
 : : | | | | | : : : | :

```

121 GKQKAYGNACRQPD 135

```

RESULT 5

RESULT 5

ORGANISM	PRELIMINARY	PRF.	144 AA.
01-OCT-2000 (1Temblarel. 15, Created)			
01-OCT-2000 (1Temblarel. 15, Last sequence update)			
01-OCT-2000 (1Temblarel. 15, Last annotation update)			
17MDA ANTIGEN (17 KOA ANTIGEN) (PRACMENTH)...			
male-killing Rickettsia from <i>Adalia bipunctata</i> .			
Bacteria: Proteobacteria; from <i>Adalia bipunctata</i> .			
Rickettsia: Proteobacteria; alpha subdivision; Rickettsiales			
Rickettsiaceae: Rickettsiae; Rickettsia			
NBL_TaxID=38028;			

Schulenbu

Puttenburger, H., J. G. V. D., Hahlf, M., Slogett, J. J., Weberley, M. K.,
 Von Land, D., Hurst, G. D. D., Maternus, M. E. N.,
 transfe evolution of male-killing: Monophyletic origin and horizontal
 generic 1st of male-killing Rickettsia (a-Proteobacteria) from two con-
 (Coleoptera: Curculionidae), Adalia bipunctata L. and A. decemmaculata L.
 Submitted (Apr. 2006, in press).
 EMBL: AJ269518; CAB6383.1; and A. decemmaculata L.
 EMBL: AJ269517; CAB6383.1; and A. decemmaculata L.
 EMBL: AJ269517; CAB6383.1; and A. decemmaculata L.

```
NON_TER      1  
NON_TER     144
```

```

SQ SEQUENCE 144 AA: 14785 MW: C8254739CA56AE7 CRC64:
Query Match 30.3%; Score 252.5; DB 2: Length 144;
Best Local Similarity 39.4%; Pred. No. 3.3e-15;
Matches 52; Conservative 24; Mismatches 45; Indels 11; Gaps 4;

OY 25 SROEVGAATGAVVGVGQLEFGKSGRVSMAIG-GAVLGGLGSKIGOSMDODK-----I 79
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
DB 17 NKCGTGTLLGAGAGALLGSGFGKKGQL-VGVGVALGALGAVLGQIGAGDEDDRRLAEL 75
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
OY 80 KLNQSLKVKAGQVTRMRNDPTGNSYSVEPRTYQRYNKOERRQCYCREFOQKAMIAIGOK 139
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
DB 76 TSGRALEAAPSQSNVEMRNDNGNMGVYTPNKTY-----RNSYGQYCREYTGTVVIGKQ 130
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
OY 140 QEITGTACPOPD 151
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
DB 131 QKAYGNACRQPD 142
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :

RESULT 6
O31065 PRELIMINARY: PRT: 137 AA.
ID O31065: Q9MW02:
AC O31065: Q9MW02:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE 17 KDA ANTIGEN (17 KDA PROTEIN) (FRAGMENT).
OS Rickettsia honei.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=37816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TT-118;
RA Billings A.N., Yu X.-J., Teel P.D., Walker D.H.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99045882; PubMed=9828442;
RA Stenos J., Roux V., Walker D., Raoult D.;
RT "Rickettsia honei sp. nov., the aetiological agent of Flinders Island
RT spotted fever in Australia.";
RL Int. J. Syst. Bacteriol. 48:1399-1404(1998).
DR EMBL: AF027124; AAB81846.1; -;
DR EMBL: AF060706; AAD20231.1; -;
DR EMBL: AF060704; AAD20230.1; -;
FT NON_TER 1
SQ SEQUENCE 137 AA: 14167 MW: 75BC1D0D745B428C CRC64:

Query Match 30.2%; Score 251.5; DB 2: Length 137;
Best Local Similarity 39.4%; Pred. No. 3.0e-15;
Matches 52; Conservative 24; Mismatches 45; Indels 11; Gaps 4;

OY 25 SROEVGAATGAVVGVGQLEFGKSGRVSMAIG-GAVLGGLGSKIGOSMDODK-----I 79
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
DB 10 NKCGTGTLLGAGAGALLGSGFGKKGQL-VGVGVALGALGAVLGQIGAGDEDDRRLAEL 68
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
OY 80 KLNQSLKVKAGQVTRMRNDPTGNSYSVEPRTYQRYNKOERRQCYCREFOQKAMIAIGOK 139
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
DB 69 TSGRALEAAPSQSNVEMRNDNGNMGVYTPNKTY-----RNSYGQYCREYTGTVVIGKQ 123
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
OY 140 QEITGTACPOPD 151
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
DB 124 QKAYGNACRQPD 135
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :

RESULT 7
O9K4M8 PRELIMINARY: PRT: 144 AA.
AC O9K4M8:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

```

```

DE 17 KDA ANTIGEN (FRAGMENT).
OS male-killing Rickettsia from Adalia decempunctata.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=120393;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20575219; PubMed=11133455;
RA Schultenburg H.J.G.V.D., Habig M., Sloggett J.J., Webberley M.K.,
RA Bertrand D., Hurst G.D.D., Majerus M.E.N.;
RT "Incidence of male-killing Rickettsia spp. (alpha-Proteobacteria) in
RT the Ten-Spot Ladybird Beetle Adalia decempunctata L. (Coleoptera:
RT coccinellidae).";
RL Appl. Environ. Microbiol. 67:270-277(2001).
DR EMBL: AJ269516; CAB96381.1; -;
FT NON_TER 1
FT NON_TER 144
SQ SEQUENCE 144 AA: 14801 MW: C825472F16A56AE7 CRC64:

Query Match 30.2%; Score 251.5; DB 2: Length 144;
Best Local Similarity 39.4%; Pred. No. 4.1e-15;
Matches 52; Conservative 24; Mismatches 45; Indels 11; Gaps 4;

OY 25 SROEVGAATGAVVGVGQLEFGKSGRVSMAIG-GAVLGGLGSKIGOSMDODK-----I 79
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
DB 17 NKCGTGTLLGAGAGALLGSGFGKKGQL-VGVGVALGALGAVLGQIGAGDEDDRRLAEL 75
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
OY 80 KLNQSLKVKAGQVTRMRNDPTGNSYSVEPRTYQRYNKOERRQCYCREFOQKAMIAIGOK 139
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
DB 76 TSGRALEAAPSQSNVEMRNDNGNMGVYTPNKTY-----RNSYGQYCREYTGTVVIGKQ 130
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
OY 140 QEITGTACPOPD 151
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
DB 131 QKAYGNACRQPD 142
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :

RESULT 8
O53154 PRELIMINARY: PRT: 154 AA.
AC O53154:
ID O53154:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE (CLONE PRB F15F 1), 5' END CDS (FRAGMENT).
OS Rickettsia sp.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=789;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93084757; PubMed=1452660;
RA Baird R.W., Lloyd M., Stenos J., Ross B.C., Stewart R.S., Dwyer B.;
RT "Characterization and comparison of Australian human spotted fever
RT group rickettsiae.";
RL J. Clin. Microbiol. 30:2896-2902(1992).
DR EMBL: M99391; AAB73386.1; -;
FT NON_TER 154
FT NON_TER 154
SQ SEQUENCE 154 AA: 15849 MW: F5C35855EDB439D2 CRC64:

Query Match 29.3%; Score 244; DB 2: Length 154;
Best Local Similarity 35.3%; Pred. No. 2.1e-14;
Matches 55; Conservative 31; Mismatches 54; Indels 16; Gaps 6;

OY 8 SLLIISV---FLVGC--AQNFSROEVAATGAVVGVGQLEFGKSGRVSMAIG-GAVL 61
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
DB 5 SKIMIIATLSMLQACNGPGMKNKGRTLLGAGAGALLGSGFGKKGQL-VGVGVALL 63
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
OY 62 GGLIGSKIGSMPODK-----IKLNQSLKVKAGQVTRMRNDPTGNSYSVEPRTYQRYN 117
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
DB 64 GAVLGQIGAGMDEDDRRLAELTSGRALEAAPSQSNVEMRNDNGNMGVYTPNKTYFNST 123
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
OY 118 KOERROQYCREFOQKAMIAIGOKQEIYGTACPOPDGR 153
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :

```

Db 124 GOD-----CRVYTOTVIGGKQOKAKYGNACROPDGO 154

RESULT 9

09F909 PRELIMINARY; PRT; 151 AA.
 AC 09F909;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE OUTR MEMBRANE PROTEIN (FRAGMENT).
 OS Rickettsia helvetica.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiellae; Rickettsia.
 NCBI_TaxID=35789;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nilsson K., Pahlson C.;
 RT "Novel peptide diagnostic reagent and kit for detection of
 RL Rickettsiosis.";
 DT Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 FT EMBL: AF181036; AAC09427.1; -
 FT NON-TER 151
 SO SEQUENCE 151 AA; 15621 MW; B77407B9C71E4B39 CRC64;

Query Match 28.4%; Score 239; DB 2; Length 151;
 Best Local Similarity 35.9%; Pred. No. 5.7e-14;

Matches 55; Conservative 28; Mismatches 54; Indels 16; Gaps 6;

OY 8 SLLIIIV---FLVGC--AQNFSROEVGATGAVGVAGOLFGKSGRVSMAIG-GAVL 61
 DB 5 SKIMIIAASMLQACNGPGMKNKQGTGLLGAGAGALLSGQFGKGGQD--VGVGALL 63
 OY 62 GGLIGSTIGSGMDQDK---IKINSLERKAKQVTRMRNPDGTGNSYSEPRRTQRYN 117
 DB 64 GAVLGQGYVAGMDEORRLAELTSQLALEAPSSGNVEMNPNNGNYGVTPMKTY----- 119
 OY 118 KQRROOYCRFPOOKAMKAGQKQEIYGTACPOP 150
 DB 120 -RNSTGYCREYTOTVIGGKQOKAKYGNACROP 151
 SO SEQUENCE 151 AA; 15621 MW; B77407B9C71E4B39 CRC64;

RESULT 10

09F001 PRELIMINARY; PRT; 131 AA.
 AC 09F001;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 17 KDA PROTEIN (FRAGMENT).
 OS Rickettsia sp. California 2.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiellae; Rickettsia.
 NCBI_TaxID=147259;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Roux V., Raoult D.;
 RT "A new SFG rickettsia isolated from fleas.";
 DT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 FT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 FT NON-TER 131
 SO SEQUENCE 131 AA; 13374 MW; 23C8B19E29FF860 CRC64;

Query Match 28.4%; Score 236.5; DB 2; Length 131;
 Best Local Similarity 38.3%; Pred. No. 8e-14;
 Matches 49; Conservative 24; Mismatches 44; Indels 11; Gaps 4;

OY 25 SROEVGATGAVGVAGOLFGKSGRVSMAIG-GAVLGGLIGSTIGSGMDQDK-----I 79
 DB 10 NKQGTGTLTGAGGALLSGQFGKGGQD--VGVGALLGAVLGQIGAGMDEDRRLAEL 68
 OY 80 KINOSLEKVAAGQVTRMRNPDGTGNSYSEPRRTQRYNKKQRRQOYCRFPOOKAMTAGOK 139
 DB 69 TSOALATPSSGTSVEMRNPDGNGHGYTPMKTY-----RNSTGYCREYTOTVIGGKQ 123
 OY 140 QEYGTAC 147
 DB 124 OKAYGNAC 131
 SO SEQUENCE 140 AA; 13344 MW; A1DCF71050DF52DF CRC64;

RESULT 11

052637 PRELIMINARY; PRT; 131 AA.
 AC 052637;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DE 17 KDA ANTIGEN (FRAGMENT).
 OS Rickettsia sp.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiellae; Rickettsia.
 NCBI_TaxID=789;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=94117373; PubMed=8288533;
 RA Herien J.H., Hurst G.D., Zhang W., Breuer J.A., Stouthamer R.,
 RT "Rickettsial relative associated with male killing in the ladybird
 RL J. Bacteriol. 176:388-394 (1994).
 DT EMBL: U04162; AAA19235.1; -
 FT NON-TER 131
 SO SEQUENCE 131 AA; 13344 MW; A1DCF71050DF52DF CRC64;

Query Match 28.4%; Score 236.5; DB 2; Length 131;
 Best Local Similarity 38.3%; Pred. No. 8e-14;
 Matches 49; Conservative 24; Mismatches 44; Indels 11; Gaps 4;

OY 25 SROEVGATGAVGVAGOLFGKSGRVSMAIG-GAVLGGLIGSTIGSGMDQDK-----I 79
 DB 10 NKQGTGTLTGAGGALLSGQFGKGGQD--VGVGALLGAVLGQIGAGMDEDRRLAEL 68
 OY 80 KINOSLEKVAAGQVTRMRNPDGTGNSYSEPRRTQRYNKKQRRQOYCRFPOOKAMTAGOK 139
 DB 69 TSOALATPSSGTSVEMRNPDGNGHGYTPMKTY-----RNSTGYCREYTOTVIGGKQ 123
 OY 140 QEYGTAC 147
 DB 124 OKAYGNAC 131
 SO SEQUENCE 140 AA; 13344 MW; A1DCF71050DF52DF CRC64;

RESULT 12

09L522 PRELIMINARY; PRT; 131 AA.
 AC 09L522;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DE 17 KDA SURFACE ANTIGEN (FRAGMENT).
 OS Rickettsia peacockii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiellae; Rickettsia.
 NCBI_TaxID=47589;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=DAE1007;
 MEDLINE=21091941; PubMed=1157215;
 RA Simser J.A., Palmer A.T., Munderloh U.G., Kurtli T.J.;

RT "Isolation of a spotted fever group rickettsia, Rickettsia peacockii,
 in a Rocky Mountain wood tick, Dermacentor andersoni, cell line."
 RT Appl. Environ. Microbiol. 67:546-552(2001).
 DR EMBL: AF260571; AAF69012.1; -.
 FT NON_TER 1
 FT NON_TER 131
 SO SEQUENCE 131 AA; 13413 MW; 228C020550CAA9D0 CRC64;

Query Match 28.3%; Score 235.5; DB 2; Length 131;
 Best Local Similarity 38.3%; Pred. No. 9.8e-14;
 Matches 49; Conservative 24; Mismatches 44; Indels 11; Gaps 4;

OY 25 SRQEVGATGAVYGVGQLFGKSGRVSMAG-GAVLGGLIGSKIGOSMDQDK---I 79
 Db 10 NKQGTGGLGAGGALLGSGQFGKKGQL-VGVGALLGAVLGQIGAGMDQDRRLAEL 68
 OY 80 KLNQSLKRVAGQVTRRNBDTGNSSYSVEPVRYQRYNKOERROOYCREFOOKAMIAQK 139
 Db 69 TSGRALETAPSGSNVWRNPDNGNYGVTPNKTY-----RNSTGQYCREVTQTVIGGKQ 123
 OY 140 QETYGTC 147
 Db 124 OKAYGNAC 131

RESULT 13
 O31208 PRELIMINARY; PRT; 105 AA.

AC 031208;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE 17 KDA ANTIGEN (FRAGMENT);
 OS Rickettsia sp. 'La Copita';
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=69475;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LA COPITA;
 RX MEDLINE=98367252; Pubmed=9701930;
 RA Billings A.N., Yu X.J., Teel P.D., Walker D.H.;
 RT "Detection of a spotted fever group rickettsia in Amblyomma cajennense
 (Acari: Ixodidae) in south Texas."
 RL J. Med. Entomol. 35:474-478(1998).
 DR EMBL: AF03499; AAB86943.1; -.
 FT NON_TER 1
 FT NON_TER 105
 SO SEQUENCE 105 AA; 11236 MW; 7BD35104701D4FB CRC64;

Query Match 26.0%; Score 216.5; DB 2; Length 105;
 Best Local Similarity 40.4%; Pred. No. 3.7e-12;
 Matches 46; Conservative 17; Mismatches 34; Indels 17; Gaps 3;

OY 42 GOLFSGKSGRVSMAGAVLGGLIGSKIGOSMDQDK---IKLNQSLKRVKAGQVTRWR 97
 Db 3 GQLVGVV-----GALLGAVLGQIGAGMDQDRRLAELTSGRALETAPSGSNVWR 54
 OY 98 NPDGNSYSVEPVRYQRYNKOERROOYCREFOOKAMIAQKQETYGTCAPDP 151
 Db 55 NPDGNGYGVTPNKTY-----RNSTGQYCREVTQTVIGGKQOKAYGNACQDP 103

RESULT 14
 O9AGC7 PRELIMINARY; PRT; 77 AA.
 AC 09AGC7;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE 17 KDA ANTIGEN (FRAGMENT);
 OS Rickettsia typhi.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Paspaliaris V., Liedtke B., Viletta L., Whiting J.L.;
 RT "Rickettsia typhus and Rickettsia felis rickettsioses in Australia."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF336794; AKK21272.1; -.
 FT NON_TER 1
 FT NON_TER 77
 SO SEQUENCE 77 AA; 7955 MW; B1E447C037263918 CRC64;

Query Match 16.7%; Score 139; DB 2; Length 77;
 Best Local Similarity 39.5%; Pred. No. 2.2e-05;
 Matches 30; Conservative 15; Mismatches 25; Indels 6; Gaps 3;

OY 42 GOLFSGKSGRVSMAG-GAVLGGLIGSKIGOSMDQDK---IKLNQSLKRVKAGQVTRWR 96
 Db 3 GSGQGHGKQGL-VGVGALLGAVLGQIGAGMDQDRRLVELTSGRALETAPSGSNVWR 61
 OY 97 RNPDTGNSYSVEPVRY 112
 Db 62 RNPDCNMGYVTPNKTY 77

RESULT 15
 O985G4 PRELIMINARY; PRT; 199 AA.

AC 0985G4;
 DT 01-OCT-2001 (Tremblrel. 18, Created)
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
 DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
 DE MLR7687 PROTEIN.
 GN MLR7687.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFE303099;
 RX MEDLINE=21082930; Pubmed=11214968;
 RA Kaneo T., Nakamura Y., Sato S., Asamiizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Igesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsuno M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti."
 RL DNA Res. 7:331-336(2000).
 DR EMBL: AP003012; BAB54098.1; -.
 KW Complete proteome.
 SO SEQUENCE 199 AA; 20517 MW; 98E6F20A734637AA CRC64;

Query Match 16.5%; Score 137; DB 16; Length 199;
 Best Local Similarity 30.9%; Pred. No. 0.00011;
 Matches 34; Conservative 16; Mismatches 50; Indels 10; Gaps 3;

OY 49 SGRVSMAGAVLGGLIGSKIGOSMDQDKIKL-----NQSLEKRVKAGQVTRRNPDGNS 104
 Db 92 SGKVTKSLISMOSGLIGSIGNGLSDEKRSALAEYKALEYTSQKYAMKGDQASHY 151
 OY 105 YSVEPVRYQRYNKOERROOYCREFOOKAMIAQKQETYGTCAPDPGRW 154
 Db 152 GEVVPAPY-RVGSQD-----CROYTQTVFTGGAGYTAGTACRNADGSW 195

Search completed: October 27, 2002, 10:59:53
 Job time : 19.0743 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:55:57 : Search time 5.83938 Seconds

(without alignments)
1067.553 Million cell updates/sec

Title: US-09-677-374-4

Perfect score: 832
Sequence: 1 MRCCLOGSSLIITVFLVGC.....ITGTACRPDGRMQVISTEK 161

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	288	34.6	159	1	17KD_RICPR
2	286	34.4	159	1	17KD_RICJA
3	285	34.3	159	1	17KD_RICCN
4	276.5	33.2	159	1	17KD_RICTY
5	268	32.2	154	1	17KD_RICAU
6	259	31.1	154	1	17KD_RICPA
7	259	31.1	154	1	17KD_RICRH
8	256	30.8	154	1	17KD_RICMO
9	255	30.6	154	1	17KD_RICAM
10	156	18.8	80	1	17KD_RICCA
11	112.5	13.5	155	1	PCP_YEREN
12	105.5	12.7	155	1	SLVB_SALTY
13	104.5	12.6	155	1	SLVB_ECOLI
14	101	12.1	155	1	PCP_HAEIN
15	100	12.0	179	1	YCFI_ECOLI
16	90	10.8	179	1	KICJ_BOVIN
17	90	10.8	1332	1	KKDO_BACSU
18	88	10.6	1585	1	YQBO_BACSU
19	86.5	10.4	72	1	OSMB_SALTY
20	85.5	10.3	72	1	OSMB_ECOLI
21	84.5	10.2	431	1	KRE2_CANAL
22	82.5	9.9	243	1	CYSW_SALTY
23	82.5	9.9	541	1	NU37_YEAST
24	82	9.9	132	1	V615_AOUAE
25	81.5	9.8	301	1	STXG_RAT
26	81.5	9.8	526	1	VPS_BTIV1
27	80	9.6	806	1	ITB7_MOUSE
28	79.5	9.6	113	1	KR33_MOUSE
29	79.5	9.6	263	1	KR33_BOVIN
30	79	9.5	593	1	KICJ_HUMAN
31	78.5	9.4	747	1	VIAD_ECOLI
32	78.5	9.4	747	1	SPD1_NEPCL
33	78	9.4	514	1	ATPA_THIFE

34	77.5	9.3	359	1	ATPA_BOVIN	P19482	bos taurus
35	77.5	9.3	467	1	HML1_MYCLE	P46724	mycobacteri
36	77.5	9.3	543	1	ATPA_RAT	P15999	rattus norv
37	77.5	9.3	553	1	ATPA_HUMAN	P25705	homo sapien
38	77	9.3	266	1	CANS_RABIT	P06813	oryctolagus
39	76.5	9.2	553	1	ATPO_BOVIN	P19483	bos taurus
40	76.5	9.2	553	1	ATPA_MOUSE	P03265	mus musc
41	76.5	9.2	569	1	KICJ_MOUSE	P02535	mus musc
42	76.5	9.2	727	1	IF2M_HUMAN	P46189	homo sapien
43	76.5	9.2	747	1	EL5_BOVIN	P04985	bos taurus
44	76.5	9.2	2516	1	CCAD_DROME	Q24270	drosophila
45	76	9.1	526	1	VPS_BTIV10	P07389	bluetongue

ALIGNMENTS

```

RESULT 1
17KD_RICPR          STANDARD;          PRT:          159 AA.
AC P16624;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 17 kda surface antigen precursor.
GN OMP OR RP833.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_Taxid=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E.
RX MEDLINE=89359171; PubMed=2768201;
RA Anderson B.E., Tzianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen
RT gene.";
RL J. Bacteriol. 171:5199-5201(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E.
RX MEDLINE=9039499; PubMed=9823893;
RA STRAIN=MADRID E.;
RA Anderson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurliand C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M28482; AAA26378.1; ALT_SEQ.
DR EMBL: AJ235273; CA15258.1; -.
DR PIR: D33971; D33971.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN_1;
KW Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.
FT SIGNAL 1
FT CHAIN 20 159
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
SQ SEQUENCE 159 AA; 16672 MW; A33D404865EBB071 CRC64;
Query Match 34.6%; Score 288; DA 1; Length 159;
Best Local Similarity 37.9%; Pred. No. 37; Indels 16; Gaps 6;
Matches 61; Conservative 32; Mismatches 52;

```

OY 8 SLLIISV---FLVGC--AQNFSROEVGAATGAVGAGQLFGKSGRVSNAIG-GAVL 61
 Db 5 SKIMIITIALASMLQACNGSSGNKOGTGTLLGAGAGALLSGQFGQKGL-VGVGALL 63
 OY 62 GGLIGSKIGOSMDODK-----IKLNOSTLEKKAQOVFRMRNPTGNSYVEEVRITYORN 117
 Db 64 GAVLGQIGAGSMDEDRRLLELTQSRLAESAPSGSNEMRNPDNGNHGYVPNKTY----- 119
 OY 118 KERRGOVCREFPOOKAMIAQKOEIYGTACPOPDGRMOVIS 158
 Db 120 -RNSAGQYCREYTOTVILGKQOKTKYGMNCRQPDGOMOVN 159

RESULT 2

17KD_RICJA
 ID 17KD_RICJA STANDARD; PRT; 159 AA.
 AC 052764;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 17 kDa surface antigen precursor.
 GN OMP.
 OS Rickettsia japonica.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OX NCBI_TaxID=35790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YH.
 RX MEDLINE=95229950; PubMed=7714214;
 RA Furuya Y., Katayama T., Yoshida Y., Kaho T.;
 RT "Specific amplification of Rickettsia japonica DNA from clinical
 RL specimens by PCR";
 CC J. Clin. Microbiol. 33:487-489(1995).
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D16515; BAA03965.1;
 CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 CC Outer membrane: Lipoprotein; Antigen; Signal.
 CC SIGNAL 1 19
 CC CHAIN 20 159 BY SIMILARITY
 CC LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
 CC SEQUENCE 159 AA; 16554 MW; CDDCE7CEBDCD6B41 CRC64;

Query Match 34.4%; Score 286; DB 1; Length 159;
 Best Local Similarity 37.9%; Pred. No. 4.6e-18;
 Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;

OY 8 SLLIISV---FLVGC--AQNFSROEVGAATGAVGAGQLFGKSGRVSNAIG-GAVL 61
 Db 5 SKIMIITIALASMLQACNGSSGNKOGTGTLLGAGAGALLSGQFGQKGL-VGVGALL 63
 OY 62 GGLIGSKIGOSMDODK-----IKLNOSTLEKKAQOVFRMRNPTGNSYVEEVRITYORN 117
 Db 64 GAVLGQIGAGSMDEDRRLLELTQSRLAESAPSGSNEMRNPDNGNHGYVPNKTY----- 119
 OY 118 KERRGOVCREFPOOKAMIAQKOEIYGTACPOPDGRMOVIS 158
 Db 120 -RNSAGQYCREYTOTVILGKQOKTKYGMNCRQPDGOMOVN 159

RESULT 3

17KD_RICCN
 ID 17KD_RICCN STANDARD; PRT; 159 AA.

AC P05372;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE 17 kDa surface antigen precursor.
 GN OMP OR RIC1287.
 OS Rickettsia conorii, and
 OS Rickettsia rickettsii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OX NCBI_TaxID=781, 783;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES-R.conorii, and R.rickettsii;
 RX MEDLINE=89359171; PubMed=2768201;
 RA Anderson B.E., Tzianabos T.;
 RT "Comparative sequence analysis of a genus-common rickettsial antigen
 RL gene";
 CC J. Bacteriol. 171:5199-5201(1989).
 CC [2]
 CC SEQUENCE FROM N.A.
 CC SPECIES-R.conorii, STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=1157893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
 CC Science 293:2093-2098(2001).
 CC [3]
 CC SEQUENCE FROM N.A.
 CC SPECIES-R.rickettsii;
 RX MEDLINE=87222152; PubMed=3108232;
 RA Anderson B.E., Regnery R.L., Carlone G.M., Tzianabos T., McDade J.E.,
 RA Fu Z.Y., Bellini W.J.;
 RT "Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia
 RL rickettsii";
 CC J. Bacteriol. 169:2385-2390(1987).
 CC [4]
 CC SEQUENCE OF 1-30 FROM N.A.
 CC SPECIES-R.rickettsii;
 RX MEDLINE=89008059; PubMed=3139629;
 RA Anderson B.E., Baumsark B.R., Bellini W.J.;
 RT "Expression of the gene encoding the 17-kilodalton antigen from
 RL Rickettsia rickettsii: transcription and posttranslational
 CC modification";
 CC J. Bacteriol. 170:4493-4500(1988).
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M28479; AAA26379.1;
 CC EMBL; M28480; AAA26376.1;
 CC EMBL; AEO08675; AAL03825.1;
 CC EMBL; M16486; AAA26381.1;
 CC EMBL; J03371; -; NOT_ANNOTATED_CDS.
 CC DR PIR; A25972; A25972.
 CC DR PIR; A31836; A31836.
 CC DR PIR; A33971; A33971.
 CC DR PIR; B33971; B33971.
 CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 CC Outer membrane: Lipoprotein; Antigen; Signal; Complete proteome.
 CC SIGNAL 1 19
 CC CHAIN 20 159
 CC LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
 CC CONFLICT 146 146 N->D (IN REF. 3).
 CC 153 153 G->E (IN REF. 3).

```

SQ SEQUENCE 159 AA; 16581 MM; 206A2BBF74FCE169 CXC64;

Query Match 34.3%; Score 285; DB 1; Length 159;
Best Local Similarity 37.9%; Pred. No. 5.6e-18;
Matches 61; Conservative 31; Mismatches 53; Indels 16; Gaps 6;

OY 8 SLLIIIV---FLVGC--AQNFSRDEVCANVGVAGQLFGKSGSRVSAIG-GAVL 61
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 5 SKIMIIATATSMLOACNGPGGNKKGSTGLLGAGGALLGSFGKGSQOL-VGVGALL 63
OY 62 GLIISIKIIGSQSMDDQDK----IKLQSLSEKYGAVGTFRNRNDTNSVSEVPRTYORYN 117
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 64 GAVLGQIGAGAGDEDDRRALIELTSORALETAPSGSNVEHRNRNDNGNYGTYTPNKTY----- 119
OY 118 KOERROGYCREFOQKAMTAGOKOELTYGTACPDGGRMOVIS 158
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 120 -RNSTGYCREYTYTVVIGGKQOKAYGNACRQPDGMQVYN 159

RESULT 4
17KD_RICTY STANDARD; PRT; 159 AA.
AC P22882;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor.
GN OMP.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=785;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89359171; PubMed=2768201;
RA Anderson B.E., Tzianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen
   gene.";
RL J. Bacteriol. 171:5199-5201(1989).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
   anchor (probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC
CC EMBL; M28481; AAA26377.1; -.
DR PIR; G33971; C33971.
DR PROSITE; PS00013; PROKAR_LIPOROTHEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT FT 159 AA; 16549 MM; 08973E2648F8DC08 CRC64;
SQ SEQUENCE 159 AA; 16549 MM; 08973E2648F8DC08 CRC64;

Query Match 33.2%; Score 276.5; DB 1; Length 159;
Best Local Similarity 39.6%; Pred. No. 3.1e-17;
Matches 55; Conservative 27; Mismatches 46; Indels 11; Gaps 4;

OY 25 SROEVCANTGANVGVAGQLFGKSGSRVSAIG-GAVLGGLIGLSIKIGSQMDQDK----I 79
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 27 NKQGTGTLGGAGGALLDSQFGRHGQOL-VGVGVALLGAVIGGIGSLDQDKRLLEL 85
OY 80 KLNQSLSEKYGAVGTFRNRNDTNSVSEVPRTYORYNKOERROGYCREFOQKAMTAGOK 139
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 86 TSGRLSEAPSSSNTEWRNPDNGNNGYTPNKTY-----RNSTGYCREYTYTVVIGGKQ 140
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 140 CEITYGTACPDGGRMOVIS 158
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

DB      141 QTVGNACRQPDGQWQVYN 159

RESULT 5
17KD_RICAU
ID      17KD_RICAU      STANDARD:      PRT:      154 AA.
AC      P50928:
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      17 kDa surface antigen precursor (Fragment).
GN      OMP.
OS      Rickettsia australis.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC      Rickettsiaceae; Rickettsiidae; Rickettsia.
CX      NCBI_TaxID=787;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Baird R.W., Ross B., Dwyer B.
RL      Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
-1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
anchor (Probable).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on
its use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC      CC
CC      EMBL: M74042; AAA26394.1;
CC      DR      PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
CC      DR      Outer membrane: Lipoprotein; Antigen; Signal.
CC      KW      SIGNAL
CC      FT      1
CC      FT      CHAIN      20      >154      17 KDA SURFACE ANTIGEN.
CC      FT      LIPID      20      20      N-ACYL DIGLYCERIDE (PROBABLE).
CC      FT      NON_TER      154      154
CC      SQ      SEQUENCE      154 AA; 15967 MW; E3AA83346FAC320 CRC64;

Query Match      32.2%; Score 268; DB 1; Length 154;
Best Local Similarity 37.8%; Pred. No. 1.6e-16;
Matches 59; Conservative 29; Mismatches 52; Indels 16; Gaps 6;

QY      8 SLLIIISV---FLVGCAG--NFSROEVGAATGAVVGVAGVGLFGSGRVSMAIG-GAVL 61
DB      5 SKIMTIALAASMLQKCNBPFGMKKGTGTLGAGAGALLGSGFGGKGL-VGVGVGALL 63
QY      62 GGLIGSKIGQSMDQDK---IKLQSLSEKVAAGVTRMRNPDTGNSYSVEPVRTYQRYN 117
DB      64 GAVLGGQAGAGMDEDDRRLAELTSGRLAETAPASGVNEMBRNDNNGYVTFPNKTYRNSN 123
QY      118 KQERQOYCRERQOKAMTAGOKOELTYGTRACPPDDR 153
DB      124 ----GQYCREYTOFVIGGKQOKAYGNACRPDQO 154

RESULT 6
17KD_RICPA
ID      17KD_RICPA      STANDARD:      PRT:      154 AA.
AC      P50930;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      17 kDa surface antigen precursor (Fragment).
GN      OMP.
OS      Rickettsia parkeri.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC      Rickettsiaceae; Rickettsiidae; Rickettsia.
CX      NCBI_TaxID=35792;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      STRAIN-MACULATUM;

```

RA Pretzman C.I., Stothard D.R., Ralph D., Clark J.B., Fuerst P.A.;
 Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC - SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 anchor (Probable).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U17008; AAB07040.1;
 KW PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 FT Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
 FT LIPID 20 20
 FT NON_TER 154 154
 SQ SEQUENCE 154 AA; 15897 MW; 5D06F45F9DBD5EEC CRC64;

Query Match 31.1%; Score 259; DB 1; Length 154;
 Best Local Similarity 36.5%; Pred. No. 9,6e-16;
 Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

OY 8 SLLIIISV---FLVGC--AQNFSRQEVGAATGAVGVAGQLFGKSGRVSMAIG-GAVL 61
 DB 5 SKIMVIALAASMLQACNGPGMKNKOGTGLLGAGALLGSGQKGGQL-VGVGVGALL 63
 OY 62 GGLIGSKIGQSDQDQK-----IKNLSLEKVKAGVTRMRNPDTGNSYSEVPTRYORYN 117
 DB 64 GAVLGQIGAGMDEQDRRLAELTSQRALETPAGSNNVWRNPUNGNYGYTPNKTY----- 119
 OY 118 KQRRQOYCREFOOKAMIAQOKOELTYGTACPOPDGR 153
 DB 120 -RNSTGQYCREYQTQVTVIGSKQOKAKAGNACLOPDQO 154

RESULT 7
 ID 17KD_RICRH STANDARD; PRT; 154 AA.
 AC P50931;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 17 kDa surface antigen precursor (Fragment).
 GN OMP.
 OS Rickettsia rhipicephali.
 CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiidae; Rickettsia.
 CC NCBI_Taxid=33992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
 Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 CC - SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 anchor (Probable).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U11020; AAB07706.1;
 KW PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 FT Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
 FT LIPID 20 20
 SQ SEQUENCE 154 AA; 15897 MW; 5D06F45F9DBD5EEC CRC64;

FT NON_TER 154 154
 SQ SEQUENCE 154 AA; 15895 MW; 0CF85AD5D96DFEFB CRC64;

Query Match 31.1%; Score 259; DB 1; Length 154;
 Best Local Similarity 36.5%; Pred. No. 9,6e-16;
 Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

OY 8 SLLIIISV---FLVGC--AQNFSRQEVGAATGAVGVAGQLFGKSGRVSMAIG-GAVL 61
 DB 5 SKIMVIALAASMLQACNGPGMKNKOGTGLLGAGALLGSGQKGGQL-VGVGVGALL 63
 OY 62 GGLIGSKIGQSDQDQK-----IKNLSLEKVKAGVTRMRNPDTGNSYSEVPTRYORYN 117
 DB 64 GAVLGQIGAGMDEQDRRLAELTSQRALETPAGSNNVWRNPUNGNYGYTPNKTY----- 119
 OY 118 KQRRQOYCREFOOKAMIAQOKOELTYGTACPOPDGR 153
 DB 120 -RNSTGQYCREYQTQVTVIGSKQOKAKAGNACLOPDQO 154

RESULT 8
 ID 17KD_RICMO STANDARD; PRT; 154 AA.
 AC P50929;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 17 kDa surface antigen precursor (Fragment).
 GN OMP.
 OS Rickettsia montana.
 CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiidae; Rickettsia.
 CC NCBI_Taxid=33991;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
 Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC - SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 anchor (Probable).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U11017; AAB07705.1;
 KW PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 FT Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
 FT LIPID 20 20
 FT NON_TER 154 154
 SQ SEQUENCE 154 AA; 15881 MW; A09C53B8769E31DA CRC64;

Query Match 30.8%; Score 256; DB 1; Length 154;
 Best Local Similarity 36.5%; Pred. No. 1,7e-15;
 Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

OY 8 SLLIIISV---FLVGC--AQNFSRQEVGAATGAVGVAGQLFGKSGRVSMAIG-GAVL 61
 DB 5 SKIMVIALAASMLQACNGPGMKNKOGTGLLGAGALLGSGQKGGQL-VGVGVGALL 63
 OY 62 GGLIGSKIGQSDQDQK-----IKNLSLEKVKAGVTRMRNPDTGNSYSEVPTRYORYN 117
 DB 64 GAVLGQIGAGMDEQDRRLAELTSQRALETPAGSNNVWRNPUNGNYGYTPNKTY----- 119
 OY 118 KQRRQOYCREFOOKAMIAQOKOELTYGTACPOPDGR 153
 DB 120 -RNSTGQYCREYQTQVTVIGSKQOKAKAGNACLOPDQO 154

```

RESULT 9
17KD_RICAM STANDARD: PRT; 154 AA.
AC P50927;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia amblyomml.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=33989;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MO 85-1084;
RA Stochard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U11013; AAB07704.1; -
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15879 MW; E4FBE4C29D943581 CRC64;

Query Match 30.6%; Score 255; DB 1; Length 154;
Best Local Similarity 36.5%; Pred. No. 2, le-15;
Matches 57; Conservative 29; Mismatches 54; Indels 16; Gaps 6;

OY 8 SSIITISV---FLVGC--AONFSROEVGAATGAVGVAGQLGKSGSRVMAIG-GAVL 61
DB 5 SKIMTIALASTLQACNGPCGMKOGTGLLGAGGALLGSGFGKGGOL-VGVGVGALL 63
OY 62 GGLIGSKIGSMDQDK---IKLNSLEKVKAGOVTRMRNPDTGNSYSEPVRTYORYN 117
DB 64 GAVLGQGVAGMGDEQDRRIARELTSOKALETAPRPSNVEWRNPNGNGYTPPKTY----- 119
OY 118 KQERRQOYCREFOQKAMIAQKQKQEIYGTACPPDGR 153
DB 120 -RNSTGQYCREYQTVYVIGGQKQKAVGNACRDPDQ 154

RESULT 10
17KD_RICAM STANDARD: PRT; 80 AA.
AC P29697;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen (Fragment).
GN OMP.
OS Rickettsia canada.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=788;
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE=92108069; Pubmed=1729713;

```

```

RA Azad A.F., Sacchi J.B. Jr., Nelson W.M., Dasch G.A.,
RA Schmidtmann E.T., Carl M.;
RT "Genetic characterization and transovarial transmission of a
RT typhus-like rickettsia found in cat fleas.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:43-46(1992).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M82879; -; NOT_ANNOTATED, CDS.
DR InterPro: IPR000437; Prok_lipoprot.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
KW Outer membrane; Lipoprotein; Antigen.
FT NON_TER 1 1
FT NON_TER 80 80
SQ SEQUENCE 80 AA; 8372 MW; AD289A48EAB19E0E CRC64;

Query Match 18.8%; Score 156; DB 1; Length 80;
Best Local Similarity 40.3%; Pred. No. 3, 9e-07;
Matches 31; Conservative 15; Mismatches 27; Indels 4; Gaps 1;

OY 42 GOLFGSGRVSNAIGAVLGLIGSKRIGSMDQDK---IKLNSLEKVKAGOVTRMR 97
DB 1 GSPFGKGGKGLIVGAGALLGATLNGIGMGDEQDRRIARELTSOKALETTPGSTIEMR 60
OY 98 NPDTGNSYSEPVRTYQ 114
DB 61 NPDNGNGYVTPSKTYK 77

RESULT 11
PCP_YEREN STANDARD: PRT; 155 AA.
AC P31484;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1996 (Rel. 37, Last annotation update)
DE Outer membrane lipoprotein pcg precursor.
GN PCP OR PCPY.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51872 / WA-C / SEROTYPE O:8;
RX MEDLINE=92121089; Pubmed=1732192;
RA Baeninger A.J., Hantke K.;
RT "A lipoprotein of Yersinia enterocolitica facilitates ferrioxamine
RT uptake in Escherichia coli.";
RL J. Bacteriol. 174:1029-1035(1992).
CC -1- FUNCTION: FACILITATES FERRIOXAMINE UPTAKE.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -1- SIMILARITY: TO E. COLI AND S. TYPHIMURUM SLTB AND TO
CC H. INFLUENZAE PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X60448; CAA42977.1; -

```

```

DR PIR: S23787, S23787.
DR HSSP: P00778, 1P04.
DR PROSITE, PS00013; PROKAR_LIPOPROTEIN, 1.
KW Outer membrane; Lipoprotein; signal.
FT SIGNAL 1
FT CHAIN 1
FT LIPID 18 15
FT LIPID 18 18
SQ SEQUENCE 155 AA; 15362 MW; 8AD6BE2132E849FA CRC64;
Query Match
Best Local Similarity 13.5%; Score 112.5; DB 1; Length 155;
Matches 37; Conservative 19; Mismatches 43; Indels 51; Gaps
QY 9 SLIIISYFLVGCQN-----FSROE-----VGAA 33
DB 7 AVAIAAVLTLCGANNNTLSCDVFSAQAKQVYTYTGTLISVPRTYIGGDDNNWGALG 66
QY 34 GAVVGAVGAGLQFGKSGKSGVSAIGAGVLTGLISKIGSGMDQDRFKL----- 66
DB 67 GAVVGELGNTVGGGTGRSLATACAGAVAGMAGQVQAGMNNTRDQVLEVRKDDGTTILV 81
QY 82 --NLSLEKAVAGQVTRRRNDPTGNSYSVSP 109
DB 127 VQKGGPTRSVQ--RVMLASSGSTVTSP 154
RESULT 12
SLYB_SALTY
ID SLYB_SALTY STANDARD; PRT; 155 AA.
AC 053549;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Outer membrane lipoprotein slyb precursor.
GN SLYB OR STM1445 OR STY1677.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
NCBI_TaxId=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S. typhimurium; STRAIN=LT2 / SCSC1412 / ATCC 700720;
RC SPECIES=S. typhimurium; MEDLINE=96133688; PubMed=8544813;
RA Ludwig A.; Tengel C.; Bauer S.; Bubert A.; Benz R.; Mollenkopf H.-J.;
RA Goebel W.;
RA "SLYB, a regulatory protein from Salmonella typhimurium, induces a
RA haemolytic and pore-forming protein in Escherichia coli."
RA Mol. Gen. Genet. 249:474-486(1995).
[2]
RP SEQUENCE FROM N.A.
RC SPECIES=S. typhimurium; STRAIN=LT2 / SCSC1412 / ATCC 700720;
RC SPECIES=S. typhimurium; MEDLINE=96133688; PubMed=8544813;
RA McClelland M.; Sanderson K.E.;
RA Courtney L.; Porwollik S.; Ali J.; Dante M.; Du F.; Hou S.; Layman D.;
RA Leonard S.; Nguyen C.; Scott K.; Holmes A.; Grewal N.; Mulvaney E.;
RA Ryan E.; Sun H.; Flores L.; Miller W.; Stoneking T.; Nhan M.;
RA Waterston R.; Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RA LT2."
RA Nature 413:852-856(2001).
[3]
RP SEQUENCE FROM N.A.
RC SPECIES=S. typhi; STRAIN=CT18;
RC SPECIES=S. typhi; MEDLINE=11677608;
RA Parkhill J.; Dougan G.; James K.D.;
RA Churcher C.; Mungall K.L.; Bentley S.D.; Thomson N.R.; Pickard D.; Main J.;
RA Baker S.; Basham D.; Brooks K.; Chillingworth T.; Conerton F.;
RA Cronin A.; Davis P.; Davies K.M.; Dowd L.; White N.; Farrar J.;
RA Feltwell T.; Hamlin N.; Haque A.; Hien T.T.; Holroyd S.; Jagers K.;
RA Krogh A.; Larsen T.S.; Leather S.; Moule S.; O'Garra P.; Parry C.;
RA Quail M.; Rutherford K.; Simmonds M.; Skellon J.; Stevens K.;
RA Whitehead S.; Barrrell B.G.;

```

```

RT      "Complete genome sequence of a multiple drug resistant Salmonella
        Enteritidis serovar Typhimurium CT18."
RL      Nature 413:848-852(2001).
CC      -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
        anchor (Potential).
CC      -1 SIMILARITY: TO E.COLI SLTB, H.INFLUENZAE PCP AND Y. ENTEROCOLITICA
        PCP.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboratio
        between the Swiss Institute of Bioinformatics and the EMBL Outstation
        at the European Bioinformatics Institute. There are no restrictions on its
        use by non-profit institutions as long as its content is in no way
        modified and this statement is not removed. Usage by and for commercia
        entities requires a license agreement (See http://www.isb.ch/announce/
        or send an email to license@isb.slb.ch).
CC      -----
DR      EMBL; S80790; AAB35871.2; -
DR      EMBL; AE008762; AA120367.1; -
DR      EMBL; AL627271; CAD01922.1; -
DR      Stryene; SG10573; SLTB.
DR      PROSITE; PS00013; PROKAR.LIPOPROTEIN; 1.
KW      Outer membrane; Lipoprotein; Signal; Complete proteome.
FT      SIGNAL
FT      CHAIN
FT      LIPID
SO      SEQUENCE
        155 AA; 15548 MW; 82APDDCDBADH55A7 CRC64;

Query Match
Best Local Similarity 24.3%; Pred. No. 0.018; Length 155;
Matches 34; Conservative 20; Mismatches 53; Indels 33; Gaps 3;

QY      1 MRGLGGSS-----IIISFLYGCAQNFSOEVGATGAVGVGA 41
Dd      15 LACCVANDLSGDVYTASAKOYNVTYGTIVRVRYTGSGSDSNVTGAIGAVLGFL 74
QY      42 GQLFGKGSRSMAIGAVLGLIGSKINGSDODDKIKLMOSLEKYKAGVTRMRNPPR 101
Dd      75 IGIIGGIGTSRIATAAGAVGAVGAGOGVSAAMKTGOVEL--EIRK-----DD 120
QY      102 GNSYSVEPVRYORYKNQER 121
Dd      121 GNTIMWYKKGCTNFPSAQOR 140

RESULT 13
SLTB_ECOLI
ID      SLTB_ECOLI
AC      P55741; P76183; STANDARD; PRT; 155 AA.
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DE      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Outer membrane lipoprotein sltb precursor.
OS      SLTB OR B1641 OR Z2835 OR ECS2350.
OS      Escherichia coli, and
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562, 83334;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=9613368; PubMed=8544813;
RA      Ludwig A., Tengel C., Bauer S., Subert A., Benz R., Molienkopf H.-J.,
RA      Goebel W.;
RT      "slta, a regulatory protein from Salmonella typhimurium, induces a
        haemolytic and pore-forming protein in Escherichia coli.";
RL      Mol. Gen. Genet. 249:474-486(1995).
RP      SEQUENCE FROM N.A.
RX      STRAIN=R12 / MG1655;
RX      MEDLINE=9742617; PubMed=9278503;
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

```

RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12:
 RX MEDLINE=97251357; Pubmed=9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
 RA Motomura K., Nakado K., Nakamura Y., Nishimoto H., Nishio Y.,
 RA Oshima T., Saito N., Sampa G., Seki Y., Sivasubraman S.,
 RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
 RA Yamamoto Y., Horiuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN (4)
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; Pubmed=11206551;
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN (5)
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RMD 0509952;
 RX MEDLINE=21156231; Pubmed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (Potential).
 CC SIMILARITY: TO S.TYPHIMURUM SLXB, H.INFLUENZAE PCP AND
 CC Y. ENTEROCOLITICA PCP.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000259; AAC74713.1; -;
 CC EMBL: D90807; BAA15402.1; -;
 CC EMBL: AE005387; AAG56630.1; -;
 CC EMBL: AP002558; BAB35773.1; -;
 CC EMBL: EG13409; SLXB.
 CC Ecogene: EG13409; SLXB.
 CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 CC Outer membrane: Lipoprotein; signal; Complete proteome.
 CC KW SIGNL 1 17 POTENTIAL.
 CC FT CHAIN 18 155 OUTER MEMBRANE LIPOPROTEIN SLXB.
 CC FT LIPID 18 18 N-ACYL DIGLYCERIDE.
 CC FT CONFLICT 98 98 A -> T (IN REF. 2).
 CC FT CONFLICT 98 98 A -> T (IN REF. 2).
 CC SQ SEQUENCE 155 AA; 15602 MW; 543EB8A4069A5FA3 CRC64;
 CC -----
 CC Query Match 12.6%; Score 104.5; DB 1; Length 155;
 CC Best Local Similarity 22.8%; Pred. No. 0.021;
 CC Matches 33; Conservative 21; Mismatches 42; Indels 49; Gaps 4;
 CC -----
 CC QY 12 TISVFLVCAQNM-----FSROE-----VGAATGAV 36
 CC DB 10 MVLGLVGCNVNDPLSGDYVTASEAKOVNVSGTIVNBPVQIOGGDDSNVIGALICGAV 69

QY 37 VGVAVAGQLFGKSGRVSMAIGAVLIGLSKIGSQMDQDKIKLNQSLKRVKAGQYTRW 96
 DB 70 LGGFLGNTVGGGTGRSLATAGAVAGVAGGCVOSAMNKTOGVEL-EIRK----- 118
 QY 97 RNPDTGNSYSVEPRTYQRYNCKER 121
 DB 119 ---DDGNTIIVVQKQGNTRFSRGR 140

 CC RESULT 14
 CC PCP_HAEIN STANDARD; PRT; 155 AA.
 CC ID PCP_HAEIN
 CC AC P10325;
 CC DT 01-MAR-1989 (Rel. 10, Created)
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Outer membrane lipoprotein PCP precursor (15 kDa lipoprotein) (PAL
 CC DE cross-reacting lipoprotein).
 CC GN PCP OR LPP OR H1579.
 CC OS Haemophilus influenzae.
 CC OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
 CC OC Haemophilus.
 CC OX NCBI_Taxid=727;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=88115138; Pubmed=2828309;
 CC RA Deich R.A., Metcalf B.J., Finn C.W., Farley J.E., Green B.A.;
 CC RT "Cloning of genes encoding a 15,000-dalton peptidoglycan-associated
 CC RT outer membrane lipoprotein and an antigenically related 15,000-dalton
 CC RT protein from Haemophilus influenzae.";
 CC RL J. Bacteriol. 170:489-496(1988).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-RD / KW20 / ATCC 51907;
 CC RX MEDLINE=9530630; Pubmed=7542800;
 CC RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 CC RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 CC RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 CC RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 CC RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 CC RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 CC RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geochagen N.S.M.,
 CC RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 CC RA Venter J.C.;
 CC RT "Whole-genome random sequencing and assembly of Haemophilus
 CC RT influenzae Rd.";
 CC RL Science 269:496-512(1995).
 CC CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor.
 CC SIMILARITY: TO E.COLI AND S.TYPHIMURUM SLXB AND TO
 CC Y. ENTEROCOLITICA PCP.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M18877; AAA24938.1; -;
 CC EMBL: U32832; AAC23228.1; -;
 CC PIR: B28543; B28543.
 CC TIGR: H11579; -;
 CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 CC Outer membrane: Lipoprotein; signal; Complete proteome.
 CC KW SIGNL 1 18
 CC FT CHAIN 19 155 OUTER MEMBRANE LIPOPROTEIN PCP.
 CC FT LIPID 19 19 N-ACYL DIGLYCERIDE.
 CC FT CONFLICT 135 143 CSLVAERFV -> VAGARVRI (IN REF. 1).
 CC FT CONFLICT 135 143 CSLVAERFV -> VAGARVRI (IN REF. 1).
 CC SQ SEQUENCE 155 AA; 15425 MW; D7880327FC0C985 CRC64;

Query Match 12.1%; Score 101; DB 1; Length 155;
 Best Local Similarity 39.6%; Pred. No. 0.043;
 Matches 21; Conservative 9; Mismatches 23; Indels 0; Gaps 0;

QY 29 VGATGAVGVAGVAGOLFGRGSGRVSMAGAVLGLIGSKIGSQSDQDKIKL 81
 DB 62 VGTGGALGAGIAGSTIGGGRGQATAAVAGALGAGIAGSKIEKMSVNGAEL 114

RESULT 15
 YCFJ_ECOLI
 ID YCFJ_ECOLI STANDARD; PRT; 179 AA.
 AC P37796; P37951;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 01-MAR-2002 (Rel. 41, Last annotation update)
 GN Hypothetical protein ycfJ.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kita-gawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horichi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [3]
 RP SEQUENCE OF 1-63 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=81236546; PubMed=6265208;
 RA Young J.G., Rogers B.L., Campbell H.D., Jaworski A., Shaw D.C.;
 RT "Nucleotide sequence coding for the respiratory NADH dehydrogenase of
 Escherichia coli. UUG initiation codon.";
 RL Eur. J. Biochem. 116:165-170(1981).
 RN [4]
 RP IDENTIFICATION.
 RX MEDLINE=95075659; PubMed=7984428;
 RA Borodovsky M., Rudd K.E., Koonin E.V.;
 RT "Intrinsic and extrinsic approaches for detecting genes in a
 bacterial genome.";
 RL Nucleic Acids Res. 22:4756-4767(1994).
 CC -1- SIMILARITY: TO RICKETTIA 17 KDA SURFACE ANTIGEN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AE000211; AAC74194.1;
 CC EMBL: D90746; BAA35925.1;
 CC EMBL: V00306; -; NOT_ANNOTATED_CDS.

DR EcoGene; EGI2444; ycfJ.
 KW Hypothetical protein; Transmembrane, Complete proteome.
 FT TRANSMEM 5 POTENTIAL.
 SQ SEQUENCE 179 AA; 18920 MW; BA5EB0DB56D45609 CRC64;

Query Match 12.0%; Score 100; DB 1; Length 179;
 Best Local Similarity 26.3%; Pred. No. 0.061;
 Matches 36; Conservative 17; Mismatches 48; Indels 36; Gaps 4;

QY 30 GAATGAVGVAGVAGOLFGRGSGRVSMAGAVLGLIGSKIGSQSDQDKIKL 89
 DB 73 GSVLAVAGVYIGHDFGGRGKDVATVAGALGCGYACNQTGSLQESD----- 120
 QY 90 AGQVTRMRNPDTGMSYSVEPYRTQRYNKOERROQYCRFQOKAMTACQKELY-----G 144
 DB 121 -----TYTTTQQRCKTYVDKSEKMLGIDVYK-----IGDQGRIRMDRDPG 162
 QY 145 TACPQPDGRMQVISTEK 161
 DB 163 TQIPL-DSNGQLINNK 178

Search completed: October 27, 2002, 10:56:30
 Job time : 6.83938 secs

B33971

C:Genetics:
A:Gene: PA1053
C:Superfamily: PAL cross-reacting lipoprotein

Query Match
Best Local Similarity 13.4%; Score 111.5; DB 2; Length 154;
Matches 35; Conservative 21; Mismatches 35; Indels 45; Gaps 4;

QY 5 LQSSLLIIS-----VFLVGC-----AQNFSROE-----
DB 1 MKKSAIIIVASFTAMALALGCGSSLTGDTYSREARVQVTRMGTIQALRPVKEIGKTP 28
QY 29 VQAATGAVGVAGVAGQLFGKSGRVSMAIGAVLGLISKIGOSMDQDQIKLNOSLEKVK 60
DB 61 IGSYAGAGVGVAGVAGVAGGKSYVAITIGAVAGLLGAMTEBLRTQCEITVREDGD 81
QY 82 --NQSLEKXKAGVTR 95
DB 121 STRAYVQVVDGQIFR 136

RESULT 11

AE0644

Probable secreted protein STY1252 [imported] - *Salmonella enterica* subsp. *enterica* serov. *Paratyphi*
C:Species: *Salmonella enterica* subsp. *enterica* serovar *Paratyphi*
A:Note: this species has also been called *Salmonella typhi*
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 14-Dec-2001
R:Accession: AE0644
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, S.; Moule, S.; Cronin, A.; Davis, P.; Davies, R.M.; Dord, L.; White, N.; Farrar, Nature 413, 848-852, 2001

A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov. *Paratyphi*
A:Reference number: AB0502; PMID:11677608
A:Accession: AE0644
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-179 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08336.1; PID:916502381; GSPDB:GN00176
C:Superfamily: PAL cross-reacting lipoprotein

Query Match
Best Local Similarity 13.1%; Score 109; DB 2; Length 179;
Matches 37; Conservative 16; Mismatches 50; Indels 32; Gaps 4;

QY 30 GAATGAVGVAGVAGQLFGKSGRVSMAIGAVLGLISKIGOSMDQDQIKLNOSLEKVK 4;
DB 73 GSVLGVAVAGVAGVAGVAGGKGRKDVATVAGLGGCYAGNQTGSMQESD----- 89
QY 90 AGCVTRMRNDPNSVSEPVRYORNRKQRRQYCCREPO--OKAMTAGCKOETITGTA 120
DB 121 -----TYYTQORCKTQVYDKSEKMLGYDVYKIGDQGRKMDKP--CTQ 146
QY 147 CPQPPGRMVOVSTK 161
DB 165 IPL-DGNQGLVANK 178

RESULT 12

887629

Hypothetical protein CC3073 [imported] - *Caulobacter crescentus*
C:Species: *Caulobacter crescentus*
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
R:Accession: 887629
R:Merman, W.C.; Feldlyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Dutkin, A.S.; Gwin, M.L.; Haft, D.H.; Kolm, Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete genome sequence of *Caulobacter crescentus*.
A:Reference number: AB7249; PMID:1173698; PMID:1129647

A:Accession: 887629
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-232 <STO>
A:Cross-references: GB:AE005673; NID:913424723; PIDN:AAK25035.1; GSPDB:GN00148
C:Genetics: CC3073

Query Match
Best Local Similarity 13.1%; Score 109; DB 2; Length 232;
Matches 32; Conservative 14; Mismatches 33; Indels 16; Gaps 2;

QY 17 LVSCAQNFSRQVGAATGAVGVAGVAGQLFGK--GSGRVSMAIGAVLGLISKIGOSM 73
DB 81 VVGKASGSKRQVGAATGAVGVAGVAGLGAAGNSLNKNGQGTGTIGAVVAGAGSLIGCKMCKSD 140
QY 74 DQDQKIKLNS-----LEKVKAGVTR 95
DB 141 AAQEVGIGYKSGFRVAQTQVAPLVKIKKMYTR 175

RESULT 13

882837

Conserved hypothetical protein XF0178 [imported] - *Xylella fastidiosa* (strain 95sc)
C:Species: *Xylella fastidiosa*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
R:Accession: 882837
R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Seq

A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A:Reference number: AB2515; PMID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: 882837
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-257 <SIM>

A:Cross-references: GB:AE003872; GB:AE003849; NID:9104975; PIDN:AAF82991.1; GSPDB:GN

A:Experimental source: strain 95sc
R:Simmons, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrero, D.M.; Carreer as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.; J.D.; Junqueira, M.L.; Kempner, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La Chado, M.A.; Madeira, A.M.B.N.; Medeira, H.M.F.; Martins, C.L.; Marques, M.V.; Martins, F.G.; Nunes, L.R.; Oliveira, M.A.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.; Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa M.; Tsuchino, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0178

Query Match
Best Local Similarity 13.1%; Score 109; DB 2; Length 257;
Matches 19; Conservative 9; Mismatches 12; Indels 0; Gaps 0;

QY 30 GAATGAVGVAGVAGQLFGKSGRVSMAIGAVLGLISKIGOSMDQDQIKLNOSLEKVK 69
DB 105 GTAIGALIGLIGVGNFGHNGKRALTAAGAVAGGFIGNEV 144

RESULT 14

AD2696

lipA protein [imported] - *Agrobacterium tumefaciens* (strain C58, Dupont)
C:Species: *Agrobacterium tumefaciens*
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
R:Accession: AD2696
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo etage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutlyavin, T.; Levy, R.; Li, M.; MCCI

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AD2696

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-142 <RUR>

A:Cross-references: GB:AE008668; PIDN:AA141986.1; PID:g17739358; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: lppA

A:Map position: circular chromosome

Query Match

12.9%; Score 107.5; DB 2; Length 142;

Best Local Similarity 27.2%; Pred. No. 0.017;

Matches 41; Conservative 20; Mismatches 61; Indels 29; Gaps 9;

QY 9 SLITISVFLVGCANFNRQEVGAATGAVVGAGOLFEGKSGRVSMAIGAVLGGIGSK 68

DB 12 SLICVSR-LSAC-----TTTGTTPAG--GSLFGR-SAOPSTPFLANLOGGIVG-K 56

QY 69 IGSMDQODKIKL-----NOSLEKVKAGQVTRWRNPDT-GNSYSVEPVRYQRYNKOERRQ 123

DB 57 SGVELDRGDOTKALEAEYKALETAIPVITWGGDDVKGVANAP---YQVGN----- 107

QY 124 QYCRFQOKAMINGOKQEIYGTACPOPDGRW 154

DB 108 QNCROYSHLTVDGRDTRVRGAACRNDGSM 138

RESULT 15

AF0289

Probable lipoprotein slyB [imported] - Yersinia pestis (strain C092)

C:Species: Yersinia pestis

C>Date: 02-NOV-2001 #sequence_revision 02-NOV-2001 #text_change 09-NOV-2001

C:Accession: AF0289

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.;

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MID:21470413; PMID:11586360

A:Accession: AF0289

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-155 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC91178.1; PID:g15980369; GSPDB:GN00175

C:Genetics:

A:Gene: slyB

C:Superfamily: PAL cross-reacting lipoprotein

Query Match

12.9%; Score 107.5; DB 2; Length 155;

Best Local Similarity 24.0%; Pred. No. 0.019;

Matches 36; Conservative 19; Mismatches 44; Indels 51; Gaps 4;

QY 9 SLITISVFLVGCANFNRQEVGAATGAVVGAGOLFEGKSGRVSMAIGAVLGGIGSK 68

DB 7 AVAIAVVTLTGCANNNTLSGDVFTASQAKOVTVSYCTLSVPRVTIQGDDNNVVGATG 66

QY 34 GAVVGGVAGOLFEGKSGRVSMAIGAVLGGIGSKIGOSMDQODKIKL----- 81

DB 67 GAVLGGFLGNAVGGTGRSLATAAGAVAGGAGVGGALNRDGVQLERKDDGQITLV 126

QY 82 --NOSLEKVKAGQVTRWRNPDTGNSYSVEP 109

DB 127 VOKGPTQPSVCGQVMLAN--SGSTITVSP 154

Search completed: October 27, 2002, 10:58:40
Job time: 11.5665 secs

THIS PAGE BLANK (USPTO)

Db 491 OGAGAAAAAAGAGGEGIRGAGGAGGCGGSGGAGGAGGAGAAAAAGAGAGCGG 550
QY 64 LIGSKIQ 71
Db 551 LGGGAGAG 558

RESULT 2

US-09-247-806-1
; Sequence 1, Application US/09247806
; Patent No. 6280747
; GENERAL INFORMATION:
; APPLICANT: PHILIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/247,806
; EARLIER FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: FR 98/01614
; NUMBER OF SEQ. ID NOS: 14
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Nephila clavipes
US-09-247-806-1

Query Match 9.4%; Score 78.5; DB 4; Length 651;
Best Local Similarity 35.3%; Pred. No. 1.7;
Matches 24; Conservative 4; Mismatches 17; Indels 23; Gaps 2;

QY 27 QEVGATCAVVG---GVAQ-----LFGSGGRVSMAGAVLGG 63
Db 491 OGAGAAAAAAGAGGEGIRGAGGAGGCGGSGGAGGAGGAGAAAAAGAGAGCGG 550
QY 64 LIGSKIQ 71
Db 551 LGGGAGAG 558

RESULT 3

US-08-425-069-2
; Sequence 2, Application US/08425069
; Patent No. 5728810
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5728810th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/425,069
; FILING DATE: 19-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-425-069-2

Query Match 9.4%; Score 78.5; DB 1; Length 718;
Best Local Similarity 35.3%; Pred. No. 1.9;
Matches 24; Conservative 4; Mismatches 17; Indels 23; Gaps 2;

QY 27 QEVGATCAVVG---GVAQ-----LFGSGGRVSMAGAVLGG 63
Db 491 OGAGAAAAAAGAGGEGIRGAGGAGGCGGSGGAGGAGGAGAAAAAGAGAGCGG 550
QY 64 LIGSKIQ 71
Db 551 LGGGAGAG 558

RESULT 4

US-08-317-844B-2
; Sequence 2, Application US/08317844B
; Patent No. 5989894
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5989894th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,844B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-317-844B-2

Query Match

9.4%; Score 78.5; DB 2; Length 718;

RESULT 6
US-09-553-498-8

Patent No. 6309861
GENERAL INFORMATION:
APPLICANT: Ambiosius, Dorothee

; APPLICANT: Schaeffner, Joerg
 ; APPLICANT: Schwarz, Elisabeth
 TITLE OF INVENTION: Process for the production of naturally folded and secreted proteins

;
; CURRENT FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: EP99107412.1

```

; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 10
SEQ ID NO 8

```

```

      : LENGTH: 255
      : TYPE: PRT
      : ORGANISM: E. coli
      :

```

US-03-553-498-8	Query Match	9.3%	Score 77.5	DB 4	Length 255;
	Best Match Similarity	30.0%	Best Wt 0.52		

5 LONGSSITTSVEFGCAONEBROEVGAATGAVV-----GGVAGOLFGKSGCBVSMATGCAV 60
 best LOCAL SIMILARITY 20.3%, FREQ. NO. 0.02;
 Matches 37; Conservative 28; Mismatches 81; Indels 31; Gaps 7;

83 LQMTSLRSEDTAMYCARDYGAV-WGOGTIVTVSSGGGGGGGGGGGGSDIELTOSPAI 141

61 LGGLGSRIGQSMDDQDKIK-LN-----QSLERKAKAGVTRWRNPDTGNISY 105
 | | | | | | | | |
142 MSASPEEVYTMTCASSSSVRYNMFOOKSGTPRKMIYDTSKSSGVAPRFGSGGSTSY 201
 | | | | | | | | |

```

QY      106 SVEPVRTYORYNKOERKOQYCHEFOOKAMT--AGOKOETYGTAQPODPGRMOVISTE 160
      1:      |      |      |      |      |      |      |      |      |      |
      2:      |      |      |      |      |      |      |      |      |      |
      3:      |      |      |      |      |      |      |      |      |      |
      4:      |      |      |      |      |      |      |      |      |      |
      5:      |      |      |      |      |      |      |      |      |      |
      6:      |      |      |      |      |      |      |      |      |      |
      7:      |      |      |      |      |      |      |      |      |      |
      8:      |      |      |      |      |      |      |      |      |      |
      9:      |      |      |      |      |      |      |      |      |      |
     10:      |      |      |      |      |      |      |      |      |      |
     11:      |      |      |      |      |      |      |      |      |      |
     12:      |      |      |      |      |      |      |      |      |      |
     13:      |      |      |      |      |      |      |      |      |      |
     14:      |      |      |      |      |      |      |      |      |      |
     15:      |      |      |      |      |      |      |      |      |      |
     16:      |      |      |      |      |      |      |      |      |      |
     17:      |      |      |      |      |      |      |      |      |      |
     18:      |      |      |      |      |      |      |      |      |      |
     19:      |      |      |      |      |      |      |      |      |      |
     20:      |      |      |      |      |      |      |      |      |      |
     21:      |      |      |      |      |      |      |      |      |      |
     22:      |      |      |      |      |      |      |      |      |      |
     23:      |      |      |      |      |      |      |      |      |      |
     24:      |      |      |      |      |      |      |      |      |      |
     25:      |      |      |      |      |      |      |      |      |      |
     26:      |      |      |      |      |      |      |      |      |      |
     27:      |      |      |      |      |      |      |      |      |      |
     28:      |      |      |      |      |      |      |      |      |      |
     29:      |      |      |      |      |      |      |      |      |      |
     30:      |      |      |      |      |      |      |      |      |      |
     31:      |      |      |      |      |      |      |      |      |      |
     32:      |      |      |      |      |      |      |      |      |      |
     33:      |      |      |      |      |      |      |      |      |      |
     34:      |      |      |      |      |      |      |      |      |      |
     35:      |      |      |      |      |      |      |      |      |      |
     36:      |      |      |      |      |      |      |      |      |      |
     37:      |      |      |      |      |      |      |      |      |      |
     38:      |      |      |      |      |      |      |      |      |      |
     39:      |      |      |      |      |      |      |      |      |      |
     40:      |      |      |      |      |      |      |      |      |      |
     41:      |      |      |      |      |      |      |      |      |      |
     42:      |      |      |      |      |      |      |      |      |      |
     43:      |      |      |      |      |      |      |      |      |      |
     44:      |      |      |      |      |      |      |      |      |      |
     45:      |      |      |      |      |      |      |      |      |      |
     46:      |      |      |      |      |      |      |      |      |      |
     47:      |      |      |      |      |      |      |      |      |      |
     48:      |      |      |      |      |      |      |      |      |      |
     49:      |      |      |      |      |      |      |      |      |      |
     50:      |      |      |      |      |      |      |      |      |      |
     51:      |      |      |      |      |      |      |      |      |      |
     52:      |      |      |      |      |      |      |      |      |      |
     53:      |      |      |      |      |      |      |      |      |      |
     54:      |      |      |      |      |      |      |      |      |      |
     55:      |      |      |      |      |      |      |      |      |      |
     56:      |      |      |      |      |      |      |      |      |      |
     57:      |      |      |      |      |      |      |      |      |      |
     58:      |      |      |      |      |      |      |      |      |      |
     59:      |      |      |      |      |      |      |      |      |      |
     60:      |      |      |      |      |      |      |      |      |      |
     61:      |      |      |      |      |      |      |      |      |      |
     62:      |      |      |      |      |      |      |      |      |      |
     63:      |      |      |      |      |      |      |      |      |      |
     64:      |      |      |      |      |      |      |      |      |      |
     65:      |      |      |      |      |      |      |      |      |      |
     66:      |      |      |      |      |      |      |      |      |      |
     67:      |      |      |      |      |      |      |      |      |      |
     68:      |      |      |      |      |      |      |      |      |      |
     69:      |      |      |      |      |      |      |      |      |      |
     70:      |      |      |      |      |      |      |      |      |      |
     71:      |      |      |      |      |      |      |      |      |      |
     72:      |      |      |      |      |      |      |      |      |      |
     73:      |      |      |      |      |      |      |      |      |      |
     74:      |      |      |      |      |      |      |      |      |      |
     75:      |      |      |      |      |      |      |      |      |      |
     76:      |      |      |      |      |      |      |      |      |      |
     77:      |      |      |      |      |      |      |      |      |      |
     78:      |      |      |      |      |      |      |      |      |      |
     79:      |      |      |      |      |      |      |      |      |      |
     80:      |      |      |      |      |      |      |      |      |      |
     81:      |      |      |      |      |      |      |      |      |      |
     82:      |      |      |      |      |      |      |      |      |      |
     83:      |      |      |      |      |      |      |      |      |      |
     84:      |      |      |      |      |      |      |      |      |      |
     85:      |      |      |      |      |      |      |      |      |      |
     86:      |      |      |      |      |      |      |      |      |      |
     87:      |      |      |      |      |      |      |      |      |      |
     88:      |      |      |      |      |      |      |      |      |      |
     89:      |      |      |      |      |      |      |      |      |      |
     90:      |      |      |      |      |      |      |      |      |      |
     91:      |      |      |      |      |      |      |      |      |      |
     92:      |      |      |      |      |      |      |      |      |      |
     93:      |      |      |      |      |      |      |      |      |      |
     94:      |      |      |      |      |      |      |      |      |      |
     95:      |      |      |      |      |      |      |      |      |      |
     96:      |      |      |      |      |      |      |      |      |      |
     97:      |      |      |      |      |      |      |      |      |      |
     98:      |      |      |      |      |      |      |      |      |      |
     99:      |      |      |      |      |      |      |      |      |      |
    100:      |      |      |      |      |      |      |      |      |      |

```

Db 202 SL----TISMEADATYCCOOWSNPLTFGAGTKLELRAAEQ-----KLISEE 249

```

RESULT /
US-09-067-351-2
: Sequence 2, Application US/09067351
      /

```

```

; Patent NO. 5994001
;
; GENERAL INFORMATION:
;
; APPLICANT: Tang, Y. Tom
;
; APPLICANT: William Tompkins

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA

```

:      COUNTRY :  USA
:      ZIP:    94304
:
:      COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

1 SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
2
3 CURRENT APPLICATION DATA:
4
5 APPLICATION NUMBER: US/09/067,351
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
103
```

FILED DATE: HEREWITH
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: GEDDONE MICHAEL C

NAME: GERNON, MICHAEL C
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0511 US
TELECOMMUNICATION INFORMATION:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 551 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: KERANOT02
US-09-067-351-2

Query Match
Best Local Similarity 9.3%; Score 77; DB 2; Length 551;
Matches 28; Conservative 28; Mismatches 53; Indels 16; Gaps 2;

Db 7 GSSLIITSVFLVCAQNFSEVGAATGAVGVAGOLFGRKSGRVSMALGAVLGGLIG 66
111 PSEFVCPGPGIOEVTVNOSLLPLHLQIDPTIQVRAERERQIKTLNKKFTSFIDKVRFL 170
114 QRYNK 118
171 EQONK 175

Db 67 SK-----IGSMDOODKIKLNQSLERKAGQVTRMRNPDTGNSYSEVPRTY 113

Db 111 PSEFVCPGPGIOEVTVNOSLLPLHLQIDPTIQVRAERERQIKTLNKKFTSFIDKVRFL 170

Db 114 QRYNK 118
171 EQONK 175

RESULT 8
US-09-360-490-2
Sequence 2, Application US/09360490
Patent No. 6221843
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Baughn, Marian
TITLE OF INVENTION: HUMAN KERATINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/360,490
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/067,351
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0511 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 551 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: KERANOT02

CLONE: 2029060
US-09-360-490-2

Query Match
Best Local Similarity 9.3%; Score 77; DB 4; Length 551;
Matches 28; Conservative 28; Mismatches 53; Indels 16; Gaps 2;

Db 7 GSSLIITSVFLVCAQNFSEVGAATGAVGVAGOLFGRKSGRVSMALGAVLGGLIG 66
54 GASFGSRSLYNLGAKRVSLNCGSSCRSGFGGRASNGFVNSG---FGYGGVGGGFGSG 110
67 SK-----IGSMDOODKIKLNQSLERKAGQVTRMRNPDTGNSYSEVPRTY 113
111 PSEFVCPGPGIOEVTVNOSLLPLHLQIDPTIQVRAERERQIKTLNKKFTSFIDKVRFL 170
114 QRYNK 118
171 EQONK 175

Db 111 PSEFVCPGPGIOEVTVNOSLLPLHLQIDPTIQVRAERERQIKTLNKKFTSFIDKVRFL 170

Db 114 QRYNK 118
171 EQONK 175

RESULT 9
US-08-374-077C-2
Sequence 2, Application US/08374077C
Patent No. 6027912
GENERAL INFORMATION:
APPLICANT: Hall, Linda M.
APPLICANT: Ren, Dejian
APPLICANT: Zheng, Wei
APPLICANT: Dubaid, Manuel Marcel Paul
TITLE OF INVENTION: Genes Encoding an Invertebrate Alpha1
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/374,077C
FILING DATE: 19-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm M.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 022650-264
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2516 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-374-077C-2

Query Match
Best Local Similarity 9.2%; Score 76.5; DB 3; Length 2516;
Matches 32; Conservative 16; Mismatches 66; Indels 13; Gaps 4;

Db 29 VGATGAVGVAGVAGOLFGRKSGRVSMALG--AVLGGLIGSKTIGS-----MDQODKIKLN 82
2389 IGSSNGSITFGSAGGLGAGSGGVG-GIGSSSIRNAFGSGSGSPSLPQHPISGTLN 2447
83 -----OSLEKVKAGQVTRMRNPDTGNSYSEVPRTYQRNRNQRYCREFQKAMIA 136

Db 2448 SPPIDNRRLRVRATVTTNNNNKSGVSONNSSLNVRANANSQNMSPGQPVQOQSPLR 2507
QY 137 GOKOEIYGT 145
Db 2508 GQGNQYSS 2516

RESULT 10

US-08-895-590-2
; Sequence 2, Application US/08895590
; Patent No. 6207410
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; APPLICANT: Ren, Dejian
; APPLICANT: Zheng, Wei
; APPLICANT: Dubald, Manuel Marcel Paul
; TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,590
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,888
; FILING DATE: 19-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm M.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 022650-263
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2516 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-895-590-2

Query Match 9.2%; Score 76.5; DB 4; Length 2516;
Best Local Similarity 24.8%; Pred. No. 17;

Matches 32; Conservative 18; Mismatches 66; Indels 13; Gaps 4;

QY 29 VGAATGAVGVAGQLFGKSGRYSAIGG--AVLGLIGSKISQ-----MDQODKIKLN 82
Db 2389 IGSSNGSIFGSGAGLGGAGSGVG-GLGSSSIRNAFGSGSGSPSLSPDHQYSGTLN 2447
QY 83 -----OSLEKVKAGVYTRMRNPDTGNSYSVEPRTYQRYNKKERRQOYCFEFOOKAMIA 136
Db 2448 SPPIDNRRLRVRATVTTNNNNKSGVSONNSSLNVRANANSQNMSPGQPVQOQSPLR 2507
QY 137 GOKOEIYGT 145
Db 2508 GQGNQYSS 2516

RESULT 11
US-07-803-633A-13
; Sequence 13, Application US/07803633A
; Patent No. 5363025

GENERAL INFORMATION:

; APPLICANT: MAZERIAN, Keyuan
; APPLICANT: LEE, Lucy F.
; APPLICANT: YANAGIDA, No. 5369025oru
; APPLICANT: OGAWA, Ryohel
; APPLICANT: LI, Yi
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VACCINE FOR
; PROTECTION AGAINST MAREK'S DISEASE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 301 No. 5369025th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/803,633A
; FILING DATE: 19911210
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1644-103P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 865 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-803-633A-13

Query Match 9.1%; Score 76; DB 1; Length 865;

Best Local Similarity 25.9%; Pred. No. 4.7; Mismatches 46; Indels 30; Gaps 8;

Matches 37; Conservative 30; Mismatches 46; Indels 30; Gaps 8;

QY 16 FLVCAQNFPSRO-EVCAATGAVGVAGQLFGKSGRYSAIGG--AVLGLIGSKISQ-----MAIGAVLGLI 65
Db 688 FMNGLAELEFNGMGVGAIGKVVVGAAGAIYSTISG-VSAFMSNPFGLAIGLIINGLV 746
QY 66 GSKIGSMDQODKIKLN-----OSLEKVKAGVYTRMRNPDTGNSYSVEPRTYQRYNKK 118
Db 747 AAFU-AVRYVNRKLKSNPKALYPMTEVLKA-QATBELHGEESDD-----LERTSI 795
QY 119 QERROQYCRE-FOOKAMIAQOK 140
Db 796 DERKLEAREMIKYMALVSABER 818

RESULT 12

US-08-220-151-6
; Sequence 6, Application US/08220151
; Patent No. 5529780
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Limbach, Keith J.
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
; CANINE HERPESVIRUS gB, gC AND gD AND USES THEREFOR
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtiss, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY

```

COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/220,151
FILING DATE: 30-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2540
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
STRANDEDNESS: Single
MOLECULE TYPE: Linear
FRAGMENT TYPE: N-terminal
US-08-220-151-6

Query Match
Best Local Similarity 23.68; Score 74.5; DB 1; Length 913;
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

QY 15 VFLVCAQNF-SROEVGATGAVGAGVAGOLFSGSRVS-----MAIGAVLGGLI 65
DB 753 VLLRGIANFFOGLDVGAAVGVKVLGATGAVISAVGAVSFLSNPGALAIGLVLGLV 812
QY 66 GSKI-----GQSDQDQIKLNOSLEKVK-----A 90
DB 813 AAFLAYHHSRLRNPKALYPVTTKLEKEDVDEGVDDEAKLDQARMIRMSIVSALE 872
QY 91 GQVTRRNPDGTGNSVSEPV-----RTYORYNKOE 120
DB 873 QQEHKARKKNSGALLASRVGAMATRRRHYQRLSESD 909

RESULT 13
US-08-413-118-6
Sequence 6, Application US/08413118
Patent No. 5688920
GENERAL INFORMATION:
APPLICANT: PAOLETTI, ENZO
APPLICANT: LIMBACH, KEITH J.
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
TITLE OF INVENTION: CANINE HERPESVIRUS 9B, 9C, AND 9D AND USES THEREFOR
NUMBER OF SEQUENCES: 128
CORRESPONDENCE ADDRESS:
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.
STREET: 530 FIFTH AVENUE, 25TH FLOOR
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/413,118
FILING DATE: 29-MAR-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: US 08/220,151
FILING DATE: 30-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: FROMMER, WILLIAM S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2670
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-0712
TELEFAX: (212) 840-3333
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
STRANDEDNESS: Single
MOLECULE TYPE: Linear
FRAGMENT TYPE: N-terminal
US-08-413-118-6

Query Match
Best Local Similarity 23.68; Score 74.5; DB 1; Length 913;
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

QY 15 VFLVCAQNF-SROEVGATGAVGAGVAGOLFSGSRVS-----MAIGAVLGGLI 65
DB 753 VLLRGIANFFOGLDVGAAVGVKVLGATGAVISAVGAVSFLSNPGALAIGLVLGLV 812
QY 66 GSKI-----GQSDQDQIKLNOSLEKVK-----A 90
DB 813 AAFLAYHHSRLRNPKALYPVTTKLEKEDVDEGVDDEAKLDQARMIRMSIVSALE 872
QY 91 GQVTRRNPDGTGNSVSEPV-----RTYORYNKOE 120
DB 873 QQEHKARKKNSGALLASRVGAMATRRRHYQRLSESD 909

RESULT 14
US-08-473-446-6
Sequence 6, Application US/08473446
Patent No. 6017542
GENERAL INFORMATION:
APPLICANT: PAOLETTI, ENZO
APPLICANT: LIMBACH, KEITH J.
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
TITLE OF INVENTION: CANINE HERPESVIRUS 9B, 9C, AND 9D AND USES THEREFOR
NUMBER OF SEQUENCES: 128
CORRESPONDENCE ADDRESS:
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.
STREET: 530 FIFTH AVENUE, 25TH FLOOR
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,446
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/413,118
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: FROMMER, WILLIAM S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2670
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-0712
TELEFAX: (212) 840-3333
INFORMATION FOR SEQ ID NO: 6:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
US-08-473-446-6

Query Match 9.0%; Score 74.5; DB 3; Length 913;
Best Local Similarity 23.6%; Pred. No. 7.4;
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

QY 15 VFLVGCANF-SROEVGATGAVGVAGOLFSGSRVS-----MAIGAVLGGLI 65
| | | | | : | | | | | : | | : | | : | | : | | :
DB 753 VLLRGIANFPGAGDVGAAVGVKVLGATGAVISAVGWSFLSNPFGALAIGLLVLAGLV 812
QY 66 GSKI-----GQSMDOODKIKLNSLEKYK-----A 90
| : | : | : | : | : | : | : | : | : | : | : | : | :
DB 813 AAFLAYRHISRLRNPMAKALYPVTTKTKEDGVDEGVDEAKLDQARDMIRYMSIVSALE 872
QY 91 GQYTRMRNPDTGNSYVEPV-----RTYQRYNKQE 120
| : | : | : | : | : | : | : | : | : | : | : | : | :
DB 873 QOEHRKARKKNSGALLASRVGAMATRRRRHYORLESED 909

RESULT 15

US-09-232-468A-2
; Sequence 2, Application US/09232468A
; Patent No. 6207165
; GENERAL INFORMATION:
; APPLICANT: AUDONNET et al.
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA AGAINST PORCINE
; FILE REFERENCE: 454313-2230
; CURRENT APPLICATION NUMBER: US/09/232,468A
; CURRENT FILING DATE: 1999-01-05
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 913
; TYPE: PRT
; ORGANISM: Pseudorabies virus
US-09-232-468A-2

Query Match 9.0%; Score 74.5; DB 4; Length 913;
Best Local Similarity 23.6%; Pred. No. 7.4;
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

QY 15 VFLVGCANF-SROEVGATGAVGVAGOLFSGSRVS-----MAIGAVLGGLI 65
| | | | | : | | | | | : | | : | | : | | : | | :
DB 753 VLLRGIANFPGAGDVGAAVGVKVLGATGAVISAVGWSFLSNPFGALAIGLLVLAGLV 812
QY 66 GSKI-----GQSMDOODKIKLNSLEKYK-----A 90
| : | : | : | : | : | : | : | : | : | : | : | : | :
DB 813 AAFLAYRHISRLRNPMAKALYPVTTKTKEDGVDEGVDEAKLDQARDMIRYMSIVSALE 872
QY 91 GQYTRMRNPDTGNSYVEPV-----RTYQRYNKQE 120
| : | : | : | : | : | : | : | : | : | : | : | : | :
DB 873 QOEHRKARKKNSGALLASRVGAMATRRRRHYORLESED 909

Search completed: October 27, 2002, 11:00:28
Job time : 9.34197 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:55:57 ; Search time 21.6891 Seconds
(without alignments)
824.509 Million cell updates/sec

Title: US-09-677-374-4

Perfect score: 832

Sequence: 1 MRCGLQSSLIITISVFLVNC.....ITGTACPDGGRWQVISTEK 161

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	832	100.0	161	22	AAB81127	Optimised Ospa pro
2	832	100.0	256	22	AAB81128	C17E2 Ospa constu
3	815	98.0	162	22	AA678025	Piscirickettsia sa
4	815	98.0	162	22	AAB81126	Ospa antigen amino
5	112	13.5	20	22	AAB81130	Ospa B-cell epitop
6	112	13.5	224	22	AAB20105	Moraxella catarrha
7	102.5	12.3	223	20	AA734487	Porphyromonas ging
8	102.5	12.3	230	20	AA734362	Porphyromonas ging
9	101	12.1	154	11	AA805799	PBOMP-2 gene prod.
10	95	11.4	309	22	ABG15906	Novel human diagno
11	88	10.6	528	22	AAB82611	Spider recombinant

12	87.5	10.5	147	22	ABB69847	Drosophila melanog
13	86.5	10.4	2309	22	ABB66232	Drosophila melanog
14	84	10.1	1251	22	ABB61254	Drosophila melanog
15	83	10.0	116	19	AA611028	H. pylori ORF 01ep
16	83	10.0	423	22	ABG30695	A fusion of anti-C
17	82	9.9	2017	22	ABG06301	Novel human diagno
18	81	9.7	566	22	ABB58019	Drosophila melanog
19	80.5	9.7	542	22	ABB65790	Drosophila melanog
20	80.5	9.7	542	22	ABB65791	Drosophila melanog
21	80.5	9.7	542	22	ABB70501	Drosophila melanog
22	79.5	9.6	618	21	AA656803	Human prostate can
23	78.5	9.4	651	20	AA440097	Spider silk protei
24	78.5	9.4	718	12	AA614308	N. clavipes draglin
25	78.5	9.4	718	19	AA653346	Nephila clavipes s
26	78.5	9.4	718	21	AA759070	N. clavipes spider
27	77.5	9.3	102	22	AA441943	Human polypeptide
28	77.5	9.3	255	21	AA611398	E. coli expression
29	77.5	9.3	255	22	AA674199	PeLB-scFvOxazoln
30	77.5	9.3	255	22	AA670769	Expression plasmid
31	77.5	9.3	255	22	AA772020	E. carotovora PeLB
32	77.5	9.3	285	20	AA741688	Human PRO284 prote
33	77.5	9.3	285	21	AA644244	Human PRO284 (UNQ2
34	77.5	9.3	285	22	AA029025	Human PRO polypept
35	77.5	9.3	285	22	AA639011	Human polypeptide
36	77.5	9.3	302	22	AA440157	Human polypeptide
37	77.5	9.3	354	22	AA119445	Human diagnostic a
38	77.5	9.3	514	22	AA036520	Pseudomonas aerugi
39	77	9.3	223	21	AA657121	Human prostate can
40	77	9.3	285	21	AA67280	Human signal pepti
41	77	9.3	285	21	AA673440	Human secreted pro
42	77	9.3	285	22	AA693288	Human protein HPI0
43	77	9.3	251	21	AA52398	Human keratin KERT
44	77	9.3	2599	21	AA75098	Neisseria meningit
45	76.5	9.2	2516	17	AA601875	Neutonal invertibr

ALIGNMENTS

RESULT 1	
AA681127	
ID	AA681127 standard; Protein; 161 AA.
XX	
AC	AA681127;
XX	
DT	11-JUL-2001 (first entry)
XX	
DE	Optimised Ospa protein 17E2 amino acid sequence.
XX	
KW	Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW	vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
KW	SRS.
XX	
OS	Piscirickettsia salmonis.
OS	Synthetic.
FH	
FT	Key
FT	Region
XX	
PN	CA2281913-A1.
XX	
PD	17-MAR-2001.
XX	
PF	17-SEP-1999; 99CA-2281913.
XX	
PR	17-SEP-1999; 99CA-2281913.
XX	
PA	(KAYW/) KAY W W.
PA	(BURI/) BURIAN J.
PA	(KUZV/) KUZV M A.
XX	
PI	Key WW, Burian J, Kuzv MA;

Location/Qualifiers
109..128
/label= B_cell_epitope

```

XX DR WPI; 2001-316844/34.
XX N-PSDB; AAF6247.
XX
XX PT Method for protecting polikilothermic fish against salmonid rickettsial
XX PT septicemia and other rickettsial diseases comprises administering a
XX PT vaccine containing the Ospa protein of Piscirickettsia salmonis
XX PS
XX PS Disclosure: Fig 5; 35pp; English.
XX
XX CC This invention relates to a method for the protection against infection
XX CC of a polikilothermic fish by the bacterial pathogen, Piscirickettsia
XX CC salmonis. The method comprises administering an immunogenic amount of a
XX CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
XX CC Ospa in the form of a vaccine. The method is used for protecting animals,
XX CC particularly polikilothermic fish, against the bacterial pathogen
XX CC P. salmonis. The method is also useful for protecting against salmonid
XX CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
XX CC sequence represents optimised P. salmonis Ospa protein 17E2. The DNA
XX CC encoding Ospa 17E2 (AAF6247) has been optimised for expression in
XX CC Escherichia coli. An Ospa protein with an N-terminal fusion partner is
XX CC used in a vaccine to create an anti-Ospa antibody response.
XX SQ
XX
XX Query Match
XX Best Local Similarity 100.0%; Score 832; DB 22; Length 161;
XX Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MRGCLQGSSLLIIISVFLVGCANFNSRQEVGAATGAVGVAGOLFEGKSGRVSMAIGAV 60
XX DB 1 MGCCLOGSSLLIIISVFLVGCANFNSRQEVGAATGAVGVAGOLFEGKSGRVSMAIGAV 60
XX QY 61 LGLIGSKIGQSMQDQKIKLNOSLEKVKAGQVTRMRNPDGTNSYSVEPVRTYQRYNQE 120
XX DB 61 LGLIGSKIGQSMQDQKIKLNOSLEKVKAGQVTRMRNPDGTNSYSVEPVRTYQRYNQE 120
XX QY 121 RROQYCREFOOKAMIAQOKOEIYGTACPOPdGRWYISTEK 161
XX DB 121 RROQYCREFOOKAMIAQOKOEIYGTACPOPdGRWYISTEK 161
XX
XX RESULT 2
XX ID AAB81128 standard; Protein; 256 AA.
XX AC AAB81128;
XX
XX DT 11-JUL-2001 (first entry)
XX
XX DE C17E2 Ospa construct with N-terminal fusion partner.
XX
XX KW Polikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
XX KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;
XX SRS; 17E2; fusion construct.
XX
XX OS Piscirickettsia salmonis.
XX OS Synthetic.
XX
XX Key
XX Location/Qualifiers
XX Region 1..95
XX /label= Undefined_N-terminal_fusion_partner
XX FT 96..256
XX /label= C17E2_Ospa
XX FT /note= "Product of Ospa gene optimised for expression in
XX FT Escherichia coli."
XX CA2281913-A1.
XX
XX PD 17-MAR-2001.
XX
XX PF 17-SEP-1999; 99CA-2281913.
XX

```

```

PR 17-SEP-1999; 99CA-2281913.
XX
XX PA (KAYW/) KAY W W.
XX PA (BURJ/) BURJAN J.
XX PA (KUZV/) KUZV M A.
XX
XX PI Kay WW, Burian J, Kuzky MA;
XX
XX DR WPI; 2001-316844/34.
XX N-PSDB; AAF6248.
XX
XX PT Method for protecting polikilothermic fish against salmonid rickettsial
XX PT septicemia and other rickettsial diseases comprises administering a
XX PT vaccine containing the Ospa protein of Piscirickettsia salmonis
XX PS
XX PS Example 4; Fig 5; 35pp; English.
XX
XX CC This invention relates to a method for the protection against infection
XX CC of a polikilothermic fish by the bacterial pathogen, Piscirickettsia
XX CC salmonis. The method comprises administering an immunogenic amount of a
XX CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
XX CC Ospa in the form of a vaccine. The method is used for protecting animals,
XX CC particularly polikilothermic fish, against the bacterial pathogen
XX CC P. salmonis. The method is also useful for protecting against salmonid
XX CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
XX CC sequence represents the amino acid sequence of C17E2, a P. salmonis Ospa
XX CC construct optimised for expression in Escherichia coli, fused to an
XX CC undefined N-terminal fusion partner. The fusion protein is used in a
XX CC vaccine to create an anti-Ospa antibody response.
XX SQ
XX
XX Query Match
XX Best Local Similarity 100.0%; Score 832; DB 22; Length 256;
XX Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MRGCLQGSSLLIIISVFLVGCANFNSRQEVGAATGAVGVAGOLFEGKSGRVSMAIGAV 60
XX DB 96 MRGCLQGSSLLIIISVFLVGCANFNSRQEVGAATGAVGVAGOLFEGKSGRVSMAIGAV 60
XX QY 61 LGLIGSKIGQSMQDQKIKLNOSLEKVKAGQVTRMRNPDGTNSYSVEPVRTYQRYNQE 155
XX DB 61 LGLIGSKIGQSMQDQKIKLNOSLEKVKAGQVTRMRNPDGTNSYSVEPVRTYQRYNQE 120
XX QY 156 LGLIGSKIGQSMQDQKIKLNOSLEKVKAGQVTRMRNPDGTNSYSVEPVRTYQRYNQE 215
XX DB 121 RROQYCREFOOKAMIAQOKOEIYGTACPOPdGRWYISTEK 161
XX QY 216 RROQYCREFOOKAMIAQOKOEIYGTACPOPdGRWYISTEK 256
XX
XX RESULT 3
XX ID AAG78025 standard; Protein; 162 AA.
XX AC AAG78025;
XX
XX DT 15-JAN-2002 (first entry)
XX
XX DE Piscirickettsia salmonis polypeptide p10.6.
XX
XX KW Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
XX KW septicemia; SRS; surface antigen; vaccine; antibacterial; fish;
XX ATCC VR-1361.
XX
XX OS Piscirickettsia salmonis.
XX
XX PN W0200168865-A2.
XX
XX PD 20-SEP-2001.
XX
XX PF 12-MAR-2001; 2001MO-GB01055.
XX
XX PR 11-MAR-2000; 2000GB-0005838.
XX PR 01-JUL-2000; 2000GB-0016080.
XX

```



```

PR 01-JUL-2000: 2000GB-0016082.
PR 29-JUL-2000: 2000GB-0018599.
XX
PA (AQUA-) AQUA HEALTH EURO LTD.
XX
PI Sismard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;
PI Burzio L;
XX
DR WPI: 2001-639050/73.
DR N-PSDB; AAF89040.
XX
PT New nucleic acids encoding an amino acid sequence homologous to the
PT surface antigen present on Piscirickettsia salmonis are useful to
PT protect fish against piscirickettsiosis.
XX
PS Claim 6; Fig 5; 25pp; English.
XX
CC The invention relates to nucleic acid sequences and the encoded protein
CC of a least part of the surface antigen present on Piscirickettsia
CC salmonis for production of a vaccine with antibacterial activity to
CC protect fish against P. salmonis which causes piscirickettsiosis, also
CC known as salmonid rickettsial septicaemia.
XX
SQ Sequence 162 AA:

Query Match 98.0%; Score 815; DB 22; Length 162;
Best Local Similarity 98.8%; Pred. No. 2e-79; Mismatches 1; Indels 0; Gaps 0;
Matches 158; Conservative 1;

QY 2 RGLGSSLLIISVFLGCAQNFROEVGAATGAVVGAVGOLFSGKSGHVSMAIGAVL 61
DB 3 RGLGSSLLIISVFLGCAQNFROEVGAATGAVVGAVGOLFSGKSGHVSMAIGAVL 62
QY 62 GGLIGSKIGSMDQDKIKLNOSLEKYKAGQVTRMRNPDGNSYSVEPRTYORYNKQER 121
DB 63 GGLIGSKIGSMDQDKIKLNOSLEKYKAGQVTRMRNPDGNSYSVEPRTYORYNKQER 122
QY 122 ROOYCREFOOKAMTAGOKEIYGTACROPDGRMOVISTEK 161
DB 123 ROOYCREFOOKAMTAGOKEIYGTACROPDGRMOVISTEK 162

RESULT 4
AAB81126
ID AAB81126 standard; Protein; 162 AA.
XX
AC AAB81126;
XX
DT 11-JUL-2001 (first entry)
XX
DE OSPA antigen amino acid sequence.
XX
KW Polkillothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OSPA; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS.
XX
OS Piscirickettsia salmonis.
XX
FH Key Location/Qualifiers
FT Region 110..129
FT /label= B_cell_epitope
XX
XX CA2281913-A1.
XX
PD 17-MAR-2001.
XX
PF 17-SEP-1999; 99CA-2281913.
XX
PR 17-SEP-1999; 99CA-2281913.
XX
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZV/) KUZV M A.

```

```

XX
PI Kay WM, Burian J, Kuzyk MA;
XX
DR WPI; 2001-316844/34.
DR N-PSDB; AAF86246.
XX
PT Method for protecting polkillothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the OSPA protein of Piscirickettsia salmonis.
XX
PS Example 2; Fig 2b; 35pp; English.
XX
CC This invention relates to a method for the protection against infection
CC of a polkillothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed OSPA, or an immunogenic fragment of
CC OSPA in the form of a vaccine. The method is used for protecting animals,
CC particularly polkillothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents P. salmonis OSPA protein. An OSPA protein with an
CC N-terminal fusion partner is used in a vaccine to create an anti-OSPA
CC antibody response.
XX
SQ Sequence 162 AA:

Query Match 98.0%; Score 815; DB 22; Length 162;
Best Local Similarity 98.8%; Pred. No. 2e-79; Mismatches 1; Indels 0; Gaps 0;
Matches 158; Conservative 1;

QY 2 RGLGSSLLIISVFLGCAQNFROEVGAATGAVVGAVGOLFSGKSGHVSMAIGAVL 61
DB 3 RGLGSSLLIISVFLGCAQNFROEVGAATGAVVGAVGOLFSGKSGHVSMAIGAVL 62
QY 62 GGLIGSKIGSMDQDKIKLNOSLEKYKAGQVTRMRNPDGNSYSVEPRTYORYNKQER 121
DB 63 GGLIGSKIGSMDQDKIKLNOSLEKYKAGQVTRMRNPDGNSYSVEPRTYORYNKQER 122
QY 122 ROOYCREFOOKAMTAGOKEIYGTACROPDGRMOVISTEK 161
DB 123 ROOYCREFOOKAMTAGOKEIYGTACROPDGRMOVISTEK 162

RESULT 5
AAB81130
ID AAB81130 standard; Peptide; 20 AA.
XX
AC AAB81130;
XX
DT 11-JUL-2001 (first entry)
XX
DE OSPA B-cell epitope peptide #2.
XX
KW Polkillothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OSPA; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS; antibody.
XX
OS Piscirickettsia salmonis.
XX
FH Key Location/Qualifiers
FT Region 110..129
FT /label= B_cell_epitope
XX
XX CA2281913-A1.
XX
PD 17-MAR-2001.
XX
PF 17-SEP-1999; 99CA-2281913.
XX
PR 17-SEP-1999; 99CA-2281913.
XX
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZV/) KUZV M A.
XX
PI Kay WM, Burian J, Kuzyk MA;

```

DR WPI: 2001-316844/34.

PT Method for protecting poikilothermic fish against salmonid rickettsial
PT septicemia and other rickettsial diseases comprises administering a
XX vaccine containing the OSPA protein of *Piscirickettsia salmonis* -
PS Example 2: Page 17, 35pp; English.

XX This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*
CC salmonis. The method comprises administering an immunogenic amount of a
CC *P. salmonis* specific antigen termed OSPA, or an immunogenic fragment of a
CC OSPA in the form of a vaccine. The method is used for protecting animals,
CC particularly poikilothermic fish, against the bacterial pathogen
CC *P. salmonis*. The method is also useful for protecting against salmonid
CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
CC sequence represents an immunogenic epitope of the *P. salmonis* OSPA
XX protein. The peptide is used to raise rabbit anti-OSPA antibodies.

SO Sequence 20 AA;

Query Match

Best Local Similarity 13.5%; Score 112; DB 22; Length 20;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 109 PVRTYQRYNKGRRQYCRE 128
Db 1 PVRTYQRYNKGRRQYCRE 20

RESULT 6

AAB20105

ID AAB20105 standard; Protein: 224 AA.

AC AAB20105;

DT 23-APR-2001 (first entry)

DE Moraxella catarrhalis BASB113 protein.

KW BASB113: infection; otitis media; pneumonia; therapy; diagnosis;
KM antibacterial; antimicrobial.

OS Moraxella catarrhalis.

PN W0200100836-A1.

PD 04-JAN-2001.

PF 23-JUN-2000; 2000MO-EP05851.

PR 25-JUN-1999; 99GB-0015044.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Thonnard J;

DR WPI: 2001-112458/12.

DR N-PSDB: AAF30043.

XX New BASB113 polypeptide isolated from Moraxella catarrhalis bacterium,
PT useful for diagnosing and producing vaccines against bacterial
PT infections such as otitis media and pneumonia
XX Claim 1: Page 67; 86pp; English.

XX The present sequence is that of BASB113 protein from Moraxella
CC catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis
CC media in children and pneumonia in adults. The invention provides
CC BASB113 polypeptides, and polynucleotides encoding them, as well as
CC expression vectors, host cells and methods for producing BASB113
CC polypeptides using recombinant methods. Also claimed is a vaccine
CC composition comprising a BASB113 polypeptide, an immunogenic

CC fragment of a BASB113 polypeptide, or a polypeptide having at least
CC 85% amino acid sequence identity to BASB113, or comprising a
CC polynucleotide encoding such a polypeptide. A claimed method of
CC diagnosing a Moraxella infection involves identifying a BASB113
CC polypeptide or antibody. A claimed therapeutic composition useful
CC in treating humans with M. catarrhalis infection comprises at least
CC 1 antibody directed against a BASB113 polypeptide. BASB113
CC polypeptides also have utility in raising specific antibodies,
CC and in screening for antibacterial drugs.

SO Sequence 224 AA;

Query Match

Best Local Similarity 13.3%; Score 111; DB 22; Length 224;

Matches 37; Conservative 15; Mismatches 35; Indels 24; Gaps 5;

OY 7 GSSLIITIVFLWGCAONFROEVCAATG-----AVGVGAQQLFGKSGRYSMATGG 58
Db 7 GVLLASSMALACGANTGT---TGNGTGGGANVNKAVTGAVAL---GCTATSKATGG 60

OY 59 -----AVLGLISKTKIGQSMDDQDKIKLNSLKVKAQGVTRMRNPDTCN 103
Db 61 EKTGRDALILGAAYGAAGAAYMERQAK-----QIEQMGQGTGVTYHDTDTGN 107

RESULT 7

AAV34487

ID AAV34487 standard; Protein: 223 AA.

AC AAV34487;

DT 25-AUG-1999 (first entry)

DE Porphyromonas gingivalis protein PG3.

KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
KM vaccine; antigenic.

OS Porphyromonas gingivalis.

PN W09929870-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98MO-AU01023.

PR 04-AUG-1998; 98AU-0005028.

PR 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

PA (CSLC-) CSL LTD.

PT Agilus CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PT Ross BC, Rothel LJ, Webb EA;

DR WPI: 1999-385613/32.

DR N-PSDB: AAX91705.

XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
XX Claim 1: Page 469; 588pp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the

OY 138 OKOE 141
DB 2231 ORAQ 2234

RESULT 14

ID ABB61254 standard; Protein; 1251 AA.

XX ABB61254;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 10554.

KW Drosophila; developmental biology; cell signalling; insecticide;

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2001; 2000US-191637P.

PS 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EM;

DR WPI: 2001-656860/75.

XX N-PSDB: ABL05357.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PS genes from Drosophila and for elucidating cell signalling and cell-cell

CC interactions -

CC Disclosure: SEQ ID NO 10554; 21pp + Sequence Listing: English.

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB16177), expressed DNA

CC sequences (AB16178-AB16179) and the encoded proteins

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1251 AA;

SO

Query Match 10.1%; Score 84; DB 22; Length 1251;

Best local Similarity 22.7%; Pred. No. 7.1;

Matches 35; Conservative 28; Mismatches 67; Indels 24; Gaps 5;

RESULT 15

ID AAY11028 standard; Protein; 116 AA.

XX AAY11028;

DT 08-JUN-1999 (first entry)

DE H. pylori ORF 01cp20708_10628177_c2-50 secreted protein.

KW Vaccine; probe; diagnostic; ORF; cell envelope protein;

OS secreted protein; cytoplasmic protein; cellular protein.

PN Helicobacter pylori.

PD WO9824475-A1.

PF 11-JUN-1998.

PR 05-DEC-1997; 97WO-US22104.

PS 14-JUL-1997; 97US-0891928.

PA 05-DEC-1996; 96US-0759625.

PI 25-MAR-1997; 97US-0823745.

DR (ASTR) ASTRA AB.

DR Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;

XX WPI: 1998-33051/29.

XX N-PSDB: AAX30557.

PT New isolated Helicobacter pylori nucleic acids - used to develop

PS products for the diagnosis, prevention and treatment of infection by

CC H. pylori and other Helicobacter species

CC Claims 37, 41; Page 190-191; 339pp; English.

CC Recombinant or substantially pure preparations of H. pylori polypeptides

CC are disclosed, together with the nucleic acids encoding them. In all,

CC 97 ORFs are shown. The proteins are variously cell envelope proteins,

CC cytoplasmic proteins, secreted proteins or other cellular proteins.

CC Vaccines containing the nucleic acids or proteins are claimed, as are

CC probes containing at least 8 nucleotides from the nucleic acid

CC sequences. The vaccines are useful for treating or reducing the risk of

CC H. pylori infections, and the probes can be used diagnostically for

CC detecting the presence of Helicobacter in a sample. The products are

CC also of use in screening for compounds having the ability to interfere

CC with the H. pylori life cycle or to inhibit H. pylori infection.

XX Sequence 116 AA;

SO

Query Match 10.0%; Score 83; DB 19; Length 116;

Best local Similarity 26.7%; Pred. No. 0.42;

Matches 31; Conservative 16; Mismatches 51; Indels 18; Gaps 4;

OY 1 MRCLGSSLLITISVFLVCAQNFNRQEVGATGAVGVAGAGLFGKSGRVSMAGAV 60

DB 11 MSGLTFCVAVLGGAMVNAVAGPKTEARGELGKRVGAVGFMVDKMGCP---VGAI 67

OY 61 LGLIGISK--IQSDODDKIKLNOSLEKRVAGVTRRNNDTGNSTSVPEVRYORYNK 116

DB 68 -GGYIGSEVGDV--EDYIRGVDRPNKPKPR-----EPINDEYD 108

Search completed: October 27, 2002, 10:57:56

Job time : 23.6891 secs

No.	Score	Match	Length	DB	ID	Description
1	815	98.0	162	2	09F9R8	09F9R8 rickettsia
2	303.5	36.5	148	2	0543R1	0543R1 rickettsia
3	283	34.0	159	2	09FP92	09FP92 rickettsia
4	258.5	31.1	137	2	0222S2	0222S2 rickettsia
5	252.5	30.3	144	2	09K2N6	09K2N6 male-R. killin
6	251.5	30.2	137	2	0310S5	0310S5 rickettsia
7	251.5	30.2	144	2	09K4W8	09K4W8 male-R. killin
8	244	29.3	154	2	053114	053114 rickettsia
9	239	28.7	151	2	09F909	09F909 rickettsia
10	236.5	28.4	131	2	09FP01	09FP01 rickettsia
11	236.5	28.4	131	2	052637	052637 rickettsia
12	235.5	28.3	131	2	09L522	09L522 rickettsia
13	216.5	26.0	105	2	031208	031208 rickettsia
14	139	16.7	77	2	09AGC7	09AGC7 rickettsia
15	137	16.5	199	16	0985G4	0985G4 rhizobium 1
16	127.5	15.3	182	16	09HX13	09HX13 pseudomonas

17	120.5	14.5	136	16	Q9Z8B9	Q9Z8B9 rhizobium m
18	111.5	13.4	152	16	Q9A1S1	Q9A1S1 pseudomonas
19	109	13.1	234	16	Q9A3X6	Q9A3X6 caulobacter
20	109	13.1	257	16	Q9PGX0	Q9PGX0 xylella fas
21	108.5	13.0	155	2	Q9F6B1	Q9F6B1 edwardsiell
22	108.5	13.0	221	16	Q9Z8T9	Q9Z8T9 rhizobium m
23	106.5	12.8	155	2	Q9RA55	Q9RA55 serratia sp
24	105	12.6	155	2	Q9RB08	Q9RB08 pectobacter
25	104	12.5	139	2	Q9Z8C4	Q9Z8C4 rhizobium 1
26	102.5	12.3	223	2	Q9XC44	Q9XC44 porphyriomon
27	100.5	12.1	153	2	Q69776	Q69776 rhizobium e
28	100.5	12.1	257	16	Q9A6M8	Q9A6M8 caulobacter
29	99.5	12.0	83	16	Q9Z1P2	Q9Z1P2 rhizobium m
30	98.5	12.0	304	16	Q9Z1P2	Q9Z1P2 pseudomonas
31	99	11.9	332	16	Q9A5X2	Q9A5X2 caulobacter
32	98	11.8	250	16	Q9CN83	Q9CN83 pasteurella
33	97	11.7	124	2	Q9AX48	Q9AX48 pseudomonas
34	97	11.5	223	16	Q9KSR1	Q9KSR1 vibrio chol
35	95.5	11.5	79	16	Q9ZMT4	Q9ZMT4 rhizobium m
36	95	11.4	172	16	P76572	P76572 escherichia
37	94	11.3	608	10	Q9SUX1	Q9SUX1 arabidopsis
38	93.5	11.2	838	2	Q9AL49	Q9AL49 shigella fl
39	93	11.2	105	16	Q98P93	Q98P93 rhizobium 1
40	91	10.9	105	16	Q983Y0	Q983Y0 rhizobium 1
41	90	10.8	137	16	Q9HH07	Q9HH07 pseudomonas
42	89.5	10.8	242	16	Q9PCD9	Q9PCD9 xylella fas
43	88	10.6	544	5	Q461T1	Q461T1 nephila cla
44	87.5	10.4	147	5	Q9VS43	Q9VS43 pseudophila
45	86.5	10.4	99	16	Q9HWP0	Q9HWP0 pseudomonas

ALIGNMENTS

RESULT 1		
Q9F9K8	PRELIMINARY:	PT: 162 AA.
ID Q9F9K8		
AC Q9F9K8;		
DT 01-MAR-2001 (TREMBLrel. 16, Created)		
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)		
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
DE 17 KDA ANTIGEN.		
GN OSPA.		
OS <i>Piscirickettsia salmonis</i> .		
OC Bacteria; Proteobacteria; gamma subdivision; <i>Piscirickettsia</i> group;		
OC <i>Piscirickettsia</i> .		
OX NCBI_TaxID=1238;		
OX [1]		
RN SEQUENCE FROM N.A.		
RP STRAIN=LF-89;		
RA Kuziy M.A., Burlan J., Thornton J.C., Kay W.W.;		
RT "Identification of a genus-common Rickettsial surface antigen in the		
RT salmonid pathogen <i>Piscirickettsia salmonis</i> .";		
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.		
RL EMBL: AF184152; AAC1/000.1; -		
QO SEQUENCE 162 AA; 17661 MW; D993E6FD9A4527E CRC64;		

Query Match 98.0%; Score 815; DB 2; Length 162;

Best Local Similarity 98.8%; Pred. No. 1.5e-65;
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0

[illegible]

RESULT 2

054381 ID 054381 PRELIMINARY; PRT: 148 AA.
 AC 054381;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE 17 KDA COMMON-ANTIGEN (FRAGMENT).
 OS Rickettsia sp.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=789;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98087556; PubMed=9425244;
 RA Davis M.J., Ying Z., Brunner B.R., Pantloja A., Ferwerda F.H.;
 RT "Rickettsial relative associated with papaya bunchy top disease."
 RL Curr. Microbiol. 36:80-84(1998).
 DR EMBL; U76907; AAC02809.1; -.
 FT NON_TER
 FT NON_TER
 SO SEQUENCE 148 AA; 15050 MW; A7AFEEDE0AE84C CRC64;

Query Match 36.5%; Score 303.5; DB 2; Length 148;
 Best Local Similarity 40.1%; Pred. No. 9.2e-20;
 Matches 55; Conservative 30; Mismatches 43; Indels 9; Gaps 2;

OY 25 SROEVGAATGAVYAGAGLFEKSGRVSMAIGAVLGLIGSKIGSDMDOK---IK 80
 DB 17 NKGGSGTLIGTLGGVGGGGRGLAAGACALLGALINGIGAGMDQRIKAEIL 76
 OY 81 LKNSLEKAVAGVTRMRNPDGNSYSVEPVRYQRNKGROOYCFRFOQKAMINGO 140
 DB 77 SORALEAPSSGVQMRNDNGNNGYVTPSKAY-----KNTGQYCREYTOYVVGKQO 131
 OY 141 ELYGTACPODGRMOVI 157
 DB 132 KAYGTACRQPDGCMQV 148

RESULT 3

09F9F2 ID 09F9F2 PRELIMINARY; PRT: 159 AA.
 AC 09F9F2;
 DT 01-MAR-2001 (TREMBLrel. 15, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 17 KDA GENUS-COMMON ANTIGEN.
 OS Rickettsia felis (Rickettsia azadi).
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=42862;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21217364; PubMed=11321078;
 RA Bouyer D.H., Stenos J., Croquet-Valdes P., Moron C.G., Popov V.L.,
 RA Zavala-Velazquez J.E., Foil L.D., Stothard D.R., Azad A.F.;
 RT "Rickettsia felis: molecular characterization of a new member of the
 RT spotted fever group."
 RL Int. J. Syst. Evol. Microbiol. 51:339-347(2001).
 DR EMBL; AF195118; AAG28452.1; -.
 SO SEQUENCE 159 AA; 16497 MW; 34C5B020AF470A1F CRC64;

Query Match 34.0%; Score 283; DB 2; Length 159;
 Best Local Similarity 37.9%; Pred. No. 6.9e-18;
 Matches 61; Conservative 30; Mismatches 54; Indels 16; Gaps 6;

OY 8 SLLIISV---FLVGC--AQNFSROEVGAATGAVYAGAGLFEKSGRVSMAIG-GAVL 61
 DB 5 SKIITIALASMLQACNGCGKMKOCTGTLGGAGAGALLGSGQFGKQQL-VGVGVALL 63

OY 62 GGLIGSKIGSDMDOK-----IKLNSLEKAVAGVTRMRNPDGNSYSVEPVRYQRN 117
 DB 64 GAVLGQIGAGMDQRIKAEILSORALEATPESGVSERNDNNGHYVTPKNTY----- 119
 OY 118 KQFRROOYCFRFOQKAMINGOKEIYGTACPODGRMOVIS 158
 DB 120 -RNSTGQYCREYTOYVVGKQOKAVGNACRQPDGLMOVN 159

RESULT 4

052252 ID 052252 PRELIMINARY; PRT: 137 AA.
 AC 052252;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 17 KDA ANTIGEN (FRAGMENT).
 OS Rickettsia coolleyi.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=69410;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX Billings A.N., Tellow G.J., Walker D.H.;
 RT "Molecular characterization of a novel spotted fever group rickettsia
 RT species from Ixodes scapularis in Texas."
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF031534; AAB95267.1; -.
 FT NON_TER
 FT NON_TER
 SO SEQUENCE 137 AA; 14215 MW; A27597A9FDB5F3C CRC64;

Query Match 31.1%; Score 258.5; DB 2; Length 137;
 Best Local Similarity 39.3%; Pred. No. 9e-16;
 Matches 53; Conservative 25; Mismatches 46; Indels 11; Gaps 4;

OY 22 ONFSROEVGAATGAVYAGAGLFEKSGRVSMAIG-GAVLGLIGSKIGSDMDOK-- 78
 DB 7 RGMNKGCTTLIGAGGALLGSGQFGKQQL-VGVGVALLGAVLGQIGAGMDQRIK 65
 OY 79 --IKLNSLEKAVAGVTRMRNPDGNSYSVEPVRYQRNKGROOYCFRFOQKAMIA 136
 DB 66 AELTSORALEAPSSGVQMRNDNGNNGYVTPKNTY-----RNSTGQYCREYTOYVIG 120
 OY 137 GQKQIYGTACPOD 151
 DB 121 GKQOKAVGNACRQPD 135

RESULT 5

09K2N6 ID 09K2N6 PRELIMINARY; PRT: 144 AA.
 AC 09K2N6;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE 17 KDA ANTIGEN (17 KDA ANTIGEN) (FRAGMENT).
 OS male-killing Rickettsia from Adalia bipunctata.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=38028;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX Schlenker H.J.G.V.D., Habig M., Sloggett J.J., Webberley M.K.,
 RA Bertrand D., Hurst G.D.D., Majerus M.E.N.;
 RT "On the evolution of male-killing: Monophyletic origin and horizontal
 RT transfer of male-killing Rickettsia (a-Proteobacteria) from two con-
 RT generic ladybirds, Adalia bipunctata L. and A. decempunctata L.
 RT (Coleoptera: Coccinellidae)."
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ269518; CAB96383.1; -.
 DR EMBL; AJ269517; CAB96382.1; -.
 FT NON_TER
 FT NON_TER
 SO SEQUENCE 144 AA; 144

Query Match	Best Local Similarity	Matches	Score	DB	Length	Indels	Gaps
Query Match	Best Local Similarity	Matches	Score	DB	Length	Indels	Gaps
25 SROEVGATGAVGACQLEFGKSGRVSMAIG-GAVLGGLGISTIGOSMDODK----	30.2%;	52;	251.5;	DB 2;	144;	11;	4;
17 NKQGTGTLGGAGGALLDSQFGKRGQL-VGVGALLGAVLGIGAGMDODRLAEL	39.4%;	52;	Pred. No. 4.1e-15;				
80 KLNQSLKLVKAGQVTRMNPDTGNSYSVEPVTRYRKNQKQERQOYCRFQOKAMIAQOK							
76 TSQRLELAPSGSNVEMRPNMGNGHYTPNKTY-----RNSTGOYCRFYOTFVIGGKO							
140 QEIVGTACPOPD							
131 QKSYGNACRQPD							
053154	PRELIMINARY;	PRT;	154 AA.				
053154							
01-NOV-1996 (TREMblrel. 01, Created)							
01-NOV-1996 (TREMblrel. 01, Last sequence update)							
01-NOV-1998 (TREMblrel. 08, Last annotation update)							
(CLONE PRB R15F 1), 5' END CDS (FRAGMENT).							
Rickettsia sp.							
Bacteria: Proteobacteria; alpha subdivision; Rickettsiales;							
Rickettsiaceae; Rickettsiidae; Rickettsia.							
NCBI_TaxID=789;							
SEQUENCE FROM N.A.							
MEDLINE=93084757; PubMed=1452660;							
Baird R.W., Lloyd M., Stenos J., Rose B.C., Stewart R.S., Dwyer B.;							
"Characterization and comparison of Australian human spotted fever							
group Rickettsiae."							
J. Clin. Microbiol. 30:2896-2902(1992).							
EMBL; M9391; AAA73386.1; -.							
NON-TER							
154							
SEQUENCE	154 AA;	15849 MM;	F5C35855DB439D2	CNC64;			
Query Match	29.3%;	Score 244;	DB 2;	Length 154;			
Best Local Similarity	35.3%;	Pred. No. 2.1e-14;					
Matches	55;	Conservative	31;	Mismatches	54;	Indels	16;
8 SLLIISV---FLVGC--AQNFSRQEVGAATGAVGVAGAGOLFQKSGSRVSMAIG-GAVL							
5 SKIMIIATATSMLOACNGCPGKNNKGTCTLLGGAGALLGDSQFGKQOL-VGVGALL							
62 GLGISKIGOSMDODK----IKLNQSLKLVKAGQVTRMNPDTGNSYSVEPVTRYRKNQKQERQOYCRFQOKAMIAQOK							
64 GAVLGQIGAGMDODRLAELTSQRALETPADSGSNVEMRPNMGNGHYTPNKTYRNST							
118 KOERQOYCRFQOKAMIAQOKQREIVGTACPOPDGR							

Db 124 G0D-----CRVYTOTVIGGKQKQKAYGNACRQPDGQ 154

RESULT 9

Q9F909 ID 09F909 PRELIMINARY; PRT; 151 AA.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
OS OUTER MEMBRANE PROTEIN (FRAGMENT).
OC Rickettsia helvetica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
RN NCBI_Taxid=35789;
RP SEQUENCE FROM N.A.
RA Nilsson K., Pahlson C.;
RT "Novel peptide diagnostic reagent and kit for detection of
Rickettsiosis."
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF181036; AAG09427.1; -.
FT NON_TER 151
SQ SEQUENCE 151 AA: 15621 MW: 877407B9C71E4B39 CRC64;

Query Match

Best Local Similarity 28.4%; Score 239; DB 2; Length 151;
Matches 55; Conservative 28; Mismatches 54; Indels 16; Gaps 6;

QY 8 SLLIISV---FLVGC--AQNRSREVGATGAVGVAGQLFGKSGRVSMAIG-GAVL 61
Db 5 SKIMITALAAMQACNGPBGKNNKGTGTLGGAGGALLSGFGKGGQL-VGVGVGALL 63
QY 62 GGLIGSKIGSMDOODK---IKLNQSLKVKAGOVTRMNPDTGNSYVEPYRYQRYN 117
Db 64 GAVLGQGVAGGDEDDRLAELTQSRALAPSSNSVEMRNPDMNGYVTPNKTY----- 119
QY 118 KOERROQYCFEFOQKAMAYGKQKQKAYGNACRQPDGQ 150
Db 120 -RNSTGYCREYTOTVIGGKQKQKAYGNACRQPDGQ 151

RESULT 10

Q9F001 ID 09F001 PRELIMINARY; PRT; 131 AA.
AC 09F001;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 17 KDA PROTEIN (FRAGMENT).
OC Rickettsia sp. California 2.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
RN NCBI_Taxid=147259;
RP SEQUENCE FROM N.A.
RA Raoult D.;
RT "A new SFG rickettsia isolated from fleas."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [12]
RP SEQUENCE FROM N.A.
RA Roux V., Raoult D.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF10693; AAC48554.1; -.
FT NON_TER 131
SQ SEQUENCE 131 AA: 13374 MW: 2308819B29FF860 CRC64;

Query Match

Best Local Similarity 28.4%; Score 236.5; DB 2; Length 131;
Matches 49; Conservative 24; Mismatches 44; Indels 11; Gaps 4;

QY 25 SROEVGATGAVGVAGQLFGKSGRVSMAIG-GAVLGIGLIGSKIGSMDOODK-----I 79
Db 10 NKQGTGTLGAGAGALGSGFGKGGQL-VGVGVGALLGAVGQIGAGMEDQDRRLAEL 68
QY 80 KLNQSLKVKAGOVTRMNPDTGNSYVEPYRYQRYNKOERROQYCFEFOQKAMAYGOK 139
Db 69 TSORALEAPSPGSGVEMRNPDMNGYVTPNKTY-----RNSTGYCREYTOTVIGGQ 123
QY 140 QEITYGTAC 147
Db 124 QKAYGNAC 131

RESULT 11

Q52637 ID 052637 PRELIMINARY; PRT; 131 AA.
AC 052637;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 17 KDA ANTIGEN (FRAGMENT).
OC Rickettsia sp.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
RN NCBI_Taxid=789;
RP SEQUENCE FROM N.A.
RA MEDLINE=94117373; PubMed=8288533;
RA Warren J.H., Hurst G.D., Zhang W., Breuer J.A., Scouthamer R.,
RT "Rickettsial relative associated with male killing in the ladybird
beetle (Adalia bipunctata)."
RL J. Bacteriol. 176:388-394(1994).
DR EMBL; U04162; AAA19235.1; -.
FT NON_TER 131
SQ SEQUENCE 131 AA: 13344 MW: A1DCFF71050DF52DF CRC64;

Query Match

Best Local Similarity 28.4%; Score 236.5; DB 2; Length 131;
Matches 49; Conservative 24; Mismatches 44; Indels 11; Gaps 4;

QY 25 SROEVGATGAVGVAGQLFGKSGRVSMAIG-GAVLGIGLIGSKIGSMDOODK-----I 79
Db 10 NKQGTGTLGAGAGALGSGFGKGGQL-VGVGVGALLGAVGQIGAGMEDQDRRLAEL 68
QY 80 KLNQSLKVKAGOVTRMNPDTGNSYVEPYRYQRYNKOERROQYCFEFOQKAMAYGOK 139
Db 69 TSORALEAPSPGSGVEMRNPDMNGYVTPNKTY-----RNSTGYCREYTOTVIGGQ 123
QY 140 QEITYGTAC 147
Db 124 QKAYGNAC 131

RESULT 12

Q9L522 ID 09L522 PRELIMINARY; PRT; 131 AA.
AC 09L522;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 17 KDA SURFACE ANTIGEN (FRAGMENT).
OC Rickettsia peacockii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
RN NCBI_Taxid=47589;
RP SEQUENCE FROM N.A.
RA STRAIN=DAE100R;
RX MEDLINE=21091941; PubMed=11157215;
RA Simser J.A., Palmer A.T., Munderloh U.G., Kurtli T.J.;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein . protein search, using sw model

Run on: October 27, 2002, 11:00:36 : Search time 6.11744 Seconds

(without alignments)
1019.028 Million cell updates/sec

Title: US-09-677-374-4

Sequence: 1 MRCCLQGSSLIITVFLVGC.....IYCTACPDPGRMQVISTEK 161

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SWISSProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	288	34.6	159	17KD_RICPR	P16624 rickettsia
2	288	34.4	159	17KD_RICJA	Q52764 rickettsia
3	285	34.3	159	17KD_RICCN	P05372 rickettsia
4	276.5	33.2	159	17KD_RICTY	P22882 rickettsia
5	268	32.2	154	17KD_RICAU	P50928 rickettsia
6	259	31.1	154	17KD_RICPA	P50930 rickettsia
7	259	31.1	154	17KD_RICRH	P50931 rickettsia
8	256	30.8	154	17KD_RICMO	P50929 rickettsia
9	255	30.6	154	17KD_RICAM	P50927 rickettsia
10	156	18.8	80	17KD_RICCA	P29697 rickettsia
11	112.5	13.5	155	PCP_YEREN	P31464 yersinia en
12	105.5	12.7	155	SLVB_SALTY	Q33349 salmonella
13	104.5	12.6	155	SLVB_ECOLI	P55741 escherichia
14	101	12.1	155	PCP_HAEIN	P10375 haemophilus
15	100	12.0	179	YCFJ_ECOLI	P37796 escherichia
16	90	10.8	526	KICJ_BOVIN	P06394 bos taurus
17	90	10.8	1332	YKDO_BACSU	P54334 bacillus su
18	88	10.6	1585	YOBQ_BACSU	P45931 bacillus su
19	86.5	10.4	72	OSMB_SALTY	P37723 salmonella
20	85.5	10.3	72	OSMB_ECOLI	P17873 escherichia
21	84.5	10.2	431	KRE2_CANAL	Q00310 candida alb
22	82.5	9.9	243	CYSH_SALTY	P17853 salmonella
23	82.5	9.9	541	NU57_YEAST	P48837 saccharomyc
24	82	9.9	132	Y615_AOUAE	O66867 aquilex aeo
25	81.5	9.8	301	STXG_RAT	Q92158 rattus norv
26	81.5	9.8	526	VP5_BTVA1	P33476 bluetongue
27	80	9.6	806	ITB7_MOUSE	P26011 mus musculu
28	79.5	9.6	113	YKR3_CAREL	P4309 caenorhabdi
29	79.5	9.6	263	CANS_BOVIN	P13135 bos taurus
30	79.5	9.5	593	KICJ_HUMAN	P13045 homo sapien
31	78.5	9.4	219	YIAD_ECOLI	P37665 escherichia
32	78.5	9.4	747	SPDI_NEPCCL	P19837 nephila cla
33	78	9.4	514	ATPA_THRIFF	P41167 thiodacillu

ALIGNMENTS

34	77.5	9.3	359	1	ATPA_BOVIN	P19482 bos taurus
35	77.5	9.3	467	1	HEMI_MYCLE	P46774 mycobacteri
36	77.5	9.3	543	1	ATPA_RAT	P15999 rattus norv
37	77.5	9.3	553	1	ATPA_HUMAN	P25705 homo sapien
38	77	9.3	266	1	CANS_RABIT	P06813 oryctolagus
39	76.5	9.2	553	1	ATPO_BOVIN	P19483 bos taurus
40	76.5	9.2	553	1	ATPA_MOUSE	O03285 mus musculu
41	76.5	9.2	569	1	KICJ_MOUSE	P02535 mus musculu
42	76.5	9.2	727	1	IF2M_HUMAN	P46189 homo sapien
43	76.5	9.2	747	1	EL5_BOVIN	P04985 bos taurus
44	76.5	9.2	2516	1	ECAD_DROME	O24270 drosophila
45	76	9.1	526	1	VP5_BTVA10	P07389 bluetongue

RESULT 1
17KD_RICPR STANDARD; PRT; 159 AA.

AC P16624;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE 17 kda surface antigen precursor.

GN OMP OR RP833.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieng; Rickettsia.
OX NCBI_Taxid=782;
[1]

RP SEQUENCE FROM N.A.
RC STRAIN-MADRID E.
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Slichter-Ponten T., Alsmark U.C.M., Pedowski R.M., Neeslund A.K.,
RT Eriksson A.S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of mitochondria."
RL Nature 396:133-140(1998).
[2] Bacteriol. 171:5199-5201(1989).
[2] Bacteriol. 171:5199-5201(1989).

RP SEQUENCE FROM N.A.
RC STRAIN-MADRID E.
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Slichter-Ponten T., Alsmark U.C.M., Pedowski R.M., Neeslund A.K.,
RT Eriksson A.S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of mitochondria."
RL Nature 396:133-140(1998).
[2] Bacteriol. 171:5199-5201(1989).
[2] Bacteriol. 171:5199-5201(1989).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: M28482; AAA26378.1; ALT_SEQ.
DR EMBL: AJ235273; CAA15258.1; -.
DR PIR: D33971; D33971.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
SQ SEQUENCE 159 AA; 16672 MW; A33D404B55EB071 CRC64;

Query Match 34.6%; Score 288; DB 1; Length 159;
Best Local Similarly 37.9%; Pred. No. 3.1e-18;
Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;

```

QY 8 SLLIITSV---FLVGC--AQNFSRQEVGATGAVVGGVAGLREKSGSRVSMAG-GAVL 61
Db 5 SKIMITALASMLQACNGSQSMNKGCTTLLGAGGLLSSQFGQGGQGV-VGVGGALL 63
QY 62 GLIISKRGSGMDQDOK-----IKNLSLEKVAQGYTRMNPDTGNSYVEPRYTRQRYN 117
Db 64 GAVLGGQIGASMDQDRLLELTSSORALESPSGNSNEMRNPNGHGVYTPKRTY----- 119
QY 118 KQRRQOYCRFQOKAMIAQKQEIYGTACPPDGRWQYIS 158
Db 120 -RNSAGQYCREYTGTVILGGKQKQKGTGNACRQPDGQMOVVN 159

```

RESULT 2

```

17KD_RICJA STANDARD: PRT: 159 AA.
AC 052764;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 17 kDa surface antigen precursor.
GN OMP.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OX Rickettsiaceae; Rickettsiidae; Rickettsia.
RN NCBI_TaxID=35790;
RP SEQUENCE FROM N.A.
RC STRAIN-YH;
RX MEDLINE=9529950; PubMed=7714214;
RA Furuya Y., Katayama T., Yoshida Y., Kaiho I.;
RT "Specific amplification of Rickettsia japonica DNA from clinical
RT specimens by PCR.";
RL J. Clin. Microbiol. 33:487-489(1995).

```

```

CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: D16515; BAA03965.1; -
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 159 N-ACYL DIGLYCERIDE (PROBABLE).
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
SQ SEQUENCE 159 AA; 16554 MW; CDDCE7CEBDCD6B41 CRC64;

```

```

Query Match 34.4%; Score 286; DB 1; Length 159;
Best Local Similarity 37.9%; Pred. No. 4; Ee -18;
Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;
QY 8 SLLIITSV---FLVGC--AQNFSRQEVGATGAVVGGVAGLREKSGSRVSMAG-GAVL 61
Db 5 SKIMITALASMLQACNGSQSMNKGCTTLLGAGGLLSSQFGQGGQGV-VGVGGALL 63
QY 62 GLIISKRGSGMDQDOK-----IKNLSLEKVAQGYTRMNPDTGNSYVEPRYTRQRYN 117
Db 64 GAVLGGQIGASMDQDRLLELTSSORALESPSGNSNEMRNPNGHGVYTPKRTY----- 119
QY 118 KQRRQOYCRFQOKAMIAQKQEIYGTACPPDGRWQYIS 158
Db 120 -RNSAGQYCREYTGTVILGGKQKQKGTGNACRQPDGQMOVVN 159

```

```

RESULT 3
17KD_RICCN STANDARD: PRT: 159 AA.
ID 17KD_RICCN

```

```

QY 8 SLLIITSV---FLVGC--AQNFSRQEVGATGAVVGGVAGLREKSGSRVSMAG-GAVL 61
Db 5 SKIMITALASMLQACNGSQSMNKGCTTLLGAGGLLSSQFGQGGQGV-VGVGGALL 63
QY 62 GLIISKRGSGMDQDOK-----IKNLSLEKVAQGYTRMNPDTGNSYVEPRYTRQRYN 117
Db 64 GAVLGGQIGASMDQDRLLELTSSORALESPSGNSNEMRNPNGHGVYTPKRTY----- 119
QY 118 KQRRQOYCRFQOKAMIAQKQEIYGTACPPDGRWQYIS 158
Db 120 -RNSAGQYCREYTGTVILGGKQKQKGTGNACRQPDGQMOVVN 159

```

```

AC P05372;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 17 kDa surface antigen precursor.
GN OMP OR RC1287.
OS Rickettsia conorii, and
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OX Rickettsiaceae; Rickettsiidae; Rickettsia.
RN NCBI_TaxID=781, 783;
RP SEQUENCE FROM N.A.
RC SPECIES=R.conorii, and R.rickettsii;
RX MEDLINE=89359171; PubMed=2768201;
RA Anderson B.E., Tzianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen
RT gene.";
RL J. Bacteriol. 171:5199-5201(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=R.conorii, STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audifren P., Fournier P.-E., Barbe V.,
RA Sanson D., Roux V., Cossart P., Weissbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=R.rickettsii;
RX MEDLINE=87222152; PubMed=3108232;
RA Anderson B.E., Regnery R.L., Carlone G.M., Tzianabos T., McDade J.E.,
RA Fu Z.Y., Bellini W.J.;
RT "Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia
RT rickettsii.";
RL J. Bacteriol. 169:2385-2390(1987).
RN [4]
RP SEQUENCE OF 1-30 FROM N.A.
RC SPECIES=R.rickettsii;
RX MEDLINE=89008059; PubMed=3139629;
RA Anderson B.E., Baumstark B.R., Bellini W.J.;
RT "Expression of the gene encoding the 17-kilodalton antigen from
RT Rickettsia rickettsii: transcription and posttranslational
RT modification.";
RL J. Bacteriol. 170:4493-4500(1988).
RN [5]
RP SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
RP anchor (Probable).
RP -----
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RP between the Swiss Institute of Bioinformatics and the EMBL outstation -
RP the European Bioinformatics Institute. There are no restrictions on its
RP use by non-profit institutions as long as its content is in no way
RP modified and this statement is not removed. Usage by and for commercial
RP entities requires a license agreement (See http://www.isb-sib.ch/announce/
RP or send an email to license@sib-sib.ch).
RP -----
DR EMBL: M28479; AAA26379.1; -
DR EMBL: M28480; AAA26376.1; -
DR EMBL: AEO08675; AAL03825.1; -
DR EMBL: M16486; AAA26381.1; -
DR EMBL: J03371; -; NOT_ANNOTATED_CDS.
DR PIR: A25972; A25972.
DR PIR: A31836; A31836.
DR PIR: A33971; A33971.
DR PIR: B33971; B33971.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 159 N-ACYL DIGLYCERIDE (PROBABLE).
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT CONFLICT 146 146 G -> E (IN REF. 3).
FT CONFLICT 153 153

```

```

Db      141 QTYGNACRPDQGMQVYN 159

RESULT 5
17KD_RICAU
ID      17KD_RICAU      STANDARD:      PRT:      154 AA.
AC      P50928:
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      17 kDa surface antigen precursor (Fragment).
GN      OMP.
OS      Rickettsia australis.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Rickettsiaceae; Rickettsiae; Rickettsia.
OX      NCBI_TaxId=787;
RN      [1]
RP      SEQUENCE FROM N.A.
RP      Baird R.W., Ross B., Dwyer B.;
RL      Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
CC      -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC      anchor (Probable).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb.sib.ch).
CC      -----
DR      EMBL; M74042; AAA26394.1; -
DR      PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW      Outer membrane; Lipoprotein; Antigen; Signal.
FT      SIGNAL 1 19 BY SIMILARITY.
FT      CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
FT      LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT      NON_TER 154 154
SQ      SEQUENCE 154 AA; 15967 MW; E3AA83346FAC320 CRC64;

Query Match 32.2%; Score 268; DB 1; Length 154;
Best Local Similarity 37.8%; Pred. No. 1.6e-16;
Matches 59; Conservative 29; Mismatches 52; Indels 16; Gaps 6;

QY      8 SLLIIIV---FLVGCAG--NFSRQEVGATGAVVGVAGQIFGKSGRVSMAIG-GAVL 61
      1 : : : : : | | | | : : : | | | | : : | : : | : : | : : | : : |
Db      5 SKIMIALAASMLQACNSPGGMKNKGCTGLLAGAGGALLGSGFGKSGQL-VGVGVALL 63
      1 : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY      62 GGLIGSKTIGOSDQODK---IKLNQSLIEKKYKAGOVYTRMRNDPTGNSIVSEVFRTYQRYN 117
      1 : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db      64 GAVLIGQIGACMDEDDERLAEITSGRALETPASGVNEMRNDNCNGYVTPINKYTRNSN 123
      1 : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY      118 KQERRQGYCRFEQOKAMTAGOKQELTYGNACPPQDR 153
      1 : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db      124 ----GQYCREYQTQTVIGRKQKAYGNACRPDQ 154
      1 : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 6
17KD_RICPA
ID      17KD_RICPA      STANDARD:      PRT:      154 AA.
AC      P50930:
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      17 kDa surface antigen precursor (Fragment).
GN      OMP.
OS      Rickettsia parkeri.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Rickettsiaceae; Rickettsiae; Rickettsia.
OX      NCBI_TaxId=35792;
RN      [1]
RP      SEQUENCE FROM N.A.
RP      STRAIN=MACULATUM.
RC

```

RA Pretzman C.I., Stothard D.R., Ralph D., Clark J.B., Fuerst P.A.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC - SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U11020; AAA82040.1;
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 >154
FT LIPID 20 20
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15897 MW; 5D06F45F9DBD5EEC CRC64;

Query Match
Best Local Similarity 31.1%; Score 259; DB 1; Length 154;
Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

QY 8 SLLIISV---FLVGC--AQNFSRQEVGATGAVGVAGOLFQKSGRVSMAIG-GAVL 61
D 5 SKIMVIALASMLQACNGPGMKNKGCTGLLGAGGALLGSGFGKSGKGL-VGVGVALL 63
QY 62 GGLIGSKIGSMDQDK-----IKLNOSLEKVKAGQVTRMRNPDTGNSYVEPVRTYQRYN 117
D 64 GAVLGQIGAGMDEQDRRLAELTSORALETPSGSNVEMRNDNGNYITPNKTY----- 119
QY 118 KQERRQOYCREFOQKAMTAGQKQEIYGTACPOPDR 153
D 120 -RNSTGYCREYQTVYVIGKQKAYGNACIQPDGQ 154

Db 120 -RNSTGYCREYQTVYVIGKQKAYGNACIQPDGQ 154

RESULT 7
17KD-RICRH STANDARD; PRT; 154 AA.
AC P50931;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia rhipicephali.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=33992;
RN [1]
RP SEQUENCE FROM N.A.
RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC - SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U11020; AAB07706.1;
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 >154
FT LIPID 20 20
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15897 MW; 5D06F45F9DBD5EEC CRC64;

Query Match
Best Local Similarity 30.8%; Score 256; DB 1; Length 154;
Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

QY 8 SLLIISV---FLVGC--AQNFSRQEVGATGAVGVAGOLFQKSGRVSMAIG-GAVL 61
D 5 SKIMVIALASMLQACNGPGMKNKGCTGLLGAGGALLGSGFGKSGKGL-VGVGVALL 63
QY 62 GGLIGSKIGSMDQDK-----IKLNOSLEKVKAGQVTRMRNPDTGNSYVEPVRTYQRYN 117
D 64 GAVLGQIGAGMDEQDRRLAELTSORALETPSGSNVEMRNDNGNYITPNKTY----- 119
QY 118 KQERRQOYCREFOQKAMTAGQKQEIYGTACPOPDR 153
D 120 -RNSTGYCREYQTVYVIGKQKAYGNACIQPDGQ 154

Db 120 -RNSTGYCREYQTVYVIGKQKAYGNACIQPDGQ 154

FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15895 MW; 0CF85AD5D96DFEFB CRC64;

Query Match
Best Local Similarity 31.1%; Score 259; DB 1; Length 154;
Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

QY 8 SLLIISV---FLVGC--AQNFSRQEVGATGAVGVAGOLFQKSGRVSMAIG-GAVL 61
D 5 SKIMVIALASMLQACNGPGMKNKGCTGLLGAGGALLGSGFGKSGKGL-VGVGVALL 63
QY 62 GGLIGSKIGSMDQDK-----IKLNOSLEKVKAGQVTRMRNPDTGNSYVEPVRTYQRYN 117
D 64 GAVLGQIGAGMDEQDRRLAELTSORALETPSGSNVEMRNDNGNYITPNKTY----- 119
QY 118 KQERRQOYCREFOQKAMTAGQKQEIYGTACPOPDR 153
D 120 -RNSTGYCREYQTVYVIGKQKAYGNACIQPDGQ 154

Db 120 -RNSTGYCREYQTVYVIGKQKAYGNACIQPDGQ 154

RESULT 8
17KD-RICMO STANDARD; PRT; 154 AA.
AC P50929;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia montana.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=33991;
RN [1]
RP SEQUENCE FROM N.A.
RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC - SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U11017; AAB07705.1;
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 >154
FT LIPID 20 20
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15881 MW; A09C53B8769E31DA CRC64;

Query Match
Best Local Similarity 30.8%; Score 256; DB 1; Length 154;
Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

QY 8 SLLIISV---FLVGC--AQNFSRQEVGATGAVGVAGOLFQKSGRVSMAIG-GAVL 61
D 5 SKIMVIALASMLQACNGPGMKNKGCTGLLGAGGALLGSGFGKSGKGL-VGVGVALL 63
QY 62 GGLIGSKIGSMDQDK-----IKLNOSLEKVKAGQVTRMRNPDTGNSYVEPVRTYQRYN 117
D 64 GAVLGQIGAGMDEQDRRLAELTSORALETPSGSNVEMRNDNGNYITPNKTY----- 119
QY 118 KQERRQOYCREFOQKAMTAGQKQEIYGTACPOPDR 153
D 120 -RNSTGYCREYQTVYVIGKQKAYGNACIQPDGQ 154

Db 120 -RNSTGYCREYQTVYVIGKQKAYGNACIQPDGQ 154


```

RESULT 9
17KD_RICAM STANDARD; PRT; 154 AA.
ID 17KD_RICAM
AC P50927;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia amblyomni.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=33989;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MO 85-1084;
RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U11013; AAB07704.1; -.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15879 MW; E4FBE4C29D943581 CRC64;

Query Match 30.6%; Score 255; DB 1; Length 154;
Best Local Similarity 36.5%; Pred. No. 2; Le-15;
Matches 57; Conservative 29; Mismatches 54; Indels 16; Gaps 6;

QY 8 SCLIIIV---FLVGC--AQNFSKQEVGAATGAVVGVAGOLGFGKSGRVSMAIG-CAVL 61
DB 5 SKIMIIAALASTLOACNGPGMKNKGTTLLGGAGALLGSGFGKGGOL-VGVGVALL 63
QY 62 GGLGSGKIGOSMDQDQK---IKLNSLEKVKACQVTRWRNPDGTGNSYSVEPVRTYGRYN 117
DB 64 GAVVGGVAGAMDEQDRIRIELTSQKALETPAPNCSNVEWRNPNGNAYGVTPNKTY---- 119
QY 118 KOERROQYCREFOQKAMIAQKOEIVGTACPPDGR 153
DB 120 -RNSTGYCREYVGTQTVVIGGKQKAVGNACRQPDGQ 154

RESULT 10
17KD_RICCA STANDARD; PRT; 80 AA.
ID 17KD_RICCA
AC P29697;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen (Fragment).
GN OMP.
OS Rickettsia canada.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=788;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92108069; PubMed=1729713;

```

```

RA Azad A.F., Sacchi J.B. Jr., Nelson W.M., Dasch G.A.,
RA Schmidtmann E.T., Carl M.;
RT "Genetic characterization and transovarial transmission of a
RT typhus-like rickettsia found in cat fleas.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:43-46(1992).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M82879; -. NOT_ANNOTATED_CDS.
DR InterPro: IPR000437; PROK_LIPOPROT.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
KW Outer membrane; Lipoprotein; Antigen.
FT NON_TER 1 1
FT NON_TER 80 80
SQ SEQUENCE 80 AA; 8372 MW; AD289A48EAB19E0E CRC64;

Query Match 18.8%; Score 156; DB 1; Length 80;
Best Local Similarity 40.3%; Pred. No. 3; 9e-07;
Matches 31; Conservative 15; Mismatches 27; Indels 4; Gaps 1;

QY 42 GOLFSGKSGRYSMAGVAGLGLGSKIGOSMDQDQK---IKLNSLEKVKACQVTRWR 97
DB 1 GSGFGKKGGLIGVAGALLGAILGNQIGAGMDQDRRLAELTSQKALETPPSTSTSEWR 60
QY 98 NPDTGNSYSVEPVRTYQ 114
DB 61 NPDNGNYGYTPSKTYK 77

RESULT 11
PCP_YEREN STANDARD; PRT; 155 AA.
ID PCP_YEREN
AC P31484;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane lipoprotein pcv precursor.
GN PCP OR PCPY.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51872 / WA-C / SEROTYPE O:8;
RX MEDLINE=92121089; PubMed=1732192;
RA Baunler A.J., Hantke K.;
RT "A lipoprotein of Yersinia enterocolitica facilitates ferrioxamine
RT uptake in Escherichia coli.";
RL J. Bacteriol. 174:1029-1035(1992).
CC -1- FUNCTION: FACILITATES FERRIOXAMINE UPTAKE.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -1- SIMILARITY: TO E. COLI AND S. TYPHIMURIO SLXB AND TO
CC H. INFLUENZAE PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X60448; CAA42977.1; -.

```

DR PIR: S23787; S23787.
 DR HSSP: P00778; 1P04.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Outer membrane; Lipoprotein; Signal.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 155 OUTER MEMBRANE LIPOPROTEIN PCP.
 FT LIPID 18 18 N-ACYL DIGLICERIDE (POTENTIAL).
 SQ SEQUENCE 155 AA; 15362 MW; 8AD6BE2132E849FA CRC64;

Query Match 13.5%; Score 112.5; DB 1; Length 155;
 Best Local Similarity 24.7%; Pred. No. 0.0044;
 Matches 37; Conservative 19; Mismatches 43; Indels 51; Gaps 4;

QY 9 SLIISVFLVGCACN-----FSRQE-----VGAT 33
 Db 7 AVAIAAATLTGACANNNTLSGDFVSAQAKOVTVTGTLISVPTTIGGDDNNYGAIG 66
 QY 34 GAVGVAGOLFGRGSRVMAIGAVLGLGSKIGSDQDKIKL----- 81
 Db 67 GAVLGGFLGTVGGGRSLATAGAGVAGMGAGVAGAMNRTDGVLEVRKDDGTTLLV 126
 QY 82 --NQSLEKVKAGVYTRWRPDTGNSYSVEP 109
 Db 127 VQKQGPTRFVSQ--RVMLASSGSTVTVP 154

RESULT 12
 SLVB_SALTY
 ID SLVB_SALTY STANDARD: PRT; 155 AA.

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 GN Outer membrane lipoprotein slvb precursor.
 GN SLVB OR STM1445 OR STY1677.
 OS Salmonella typhimurium, and
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 ON NCBI_TaxID=602, 601;
 RX SPECIES-S.typhimurium;
 RX MEDLINE=96133688; PubMed=8544813;
 RA Ludwig A., Tengel C., Bauer S., Hubert A., Benz R., Mollenkopf H.-J.,
 RA Goebel W.;
 RT "Slvb, a regulatory protein from Salmonella typhimurium, induces a
 RT haemolytic and pore-forming protein in Escherichia coli.";
 RL Mol. Gen. Genet. 249:474-486(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S.typhimurium; STRAIN-LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534947; PubMed=11677608;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S.typhi; STRAIN-CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Cronin A., Davis P., Brooks K., Chillingworth T., Connor P.,
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jorgels K.,
 RA Krogh A., Larsen T.S., Leather A., Hien T.T., Holroyd S., Jorgels K.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrell B.G.;

RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (Potential).
 CC -1- SIMILARITY: TO E.COLI SLVB, H.INFLUENZAE PCP AND Y. ENTEROCOLITICA
 CC PCP.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: S80790; AMB35871.2; -
 DR EMBL: AE008762; AL20367.1; -
 DR EMBL: AL627271; CAD01922.1; -
 DR StyGene: SG10573; slvb.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 DR Outer membrane; Lipoprotein; Signal; Complete proteome.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 155 OUTER MEMBRANE LIPOPROTEIN SLVB.
 FT LIPID 18 18 N-ACYL DIGLICERIDE.
 SQ SEQUENCE 155 AA; 15548 MW; 82FDCCDBAD55A7 CRC64;

Query Match 12.7%; Score 105.5; DB 1; Length 155;
 Best Local Similarity 24.3%; Pred. No. 0.018;
 Matches 34; Conservative 20; Mismatches 53; Indels 33; Gaps 3;

QY 1 MRCLGSSF-----IIISVFLVGCACNFSRQEVGAATGAVGVA 41
 Db 15 LACGVNNDLSGDDVYASPAKQVNTTGTIVNAPVIOGDDSNVIGAGVAGLGL 74
 QY 42 GOLFGSGSRVMAIGAVLGLGSKIGSDQDKIKLQNSLEKVKAGVTRWRPDT 101
 Db 75 GNTIGGTRSLATAGAGVAGMGAGVAGAMNRTDGVLEVRKDDGTTLLV 126
 QY 102 GNSYSVEPRTYRYNKOER 121
 Db 121 GNTIMVQKQNTFRSAGQR 140

RESULT 13

SLVB_ECOLI
 ID SLVB_ECOLI STANDARD: PRT; 155 AA.

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 GN Outer membrane lipoprotein slvb precursor.
 GN SLVB OR BL641 OR Z2655 OR ECS2350.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 ON NCBI_TaxID=562, 83334;
 RX SPECIES FROM N.A.
 RX MEDLINE=96133688; PubMed=8544813;
 RA Ludwig A., Tengel C., Bauer S., Hubert A., Benz R., Mollenkopf H.-J.,
 RA Goebel W.;
 RT "Slvb, a regulatory protein from Salmonella typhimurium, induces a
 RT haemolytic and pore-forming protein in Escherichia coli.";
 RL Mol. Gen. Genet. 249:474-486(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MC1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 11:05:56 ; Search time 10.2884 Seconds
(without alignments)
1503.668 Million cell updates/sec

Title: US-09-677-374-4

Perfect score: 832

Sequence: 1 MRGCGSSSLIISVFLVGC.....IYGTACPDGDMQVISTEK 161

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	288	34.6	159	2 D33971	ricketsial common
2	285	34.3	159	2 B33971	ricketsial common
3	285	34.3	159	2 A33971	ricketsial common
4	285	34.3	159	2 G97860	17K surface antigen
5	276.5	33.2	159	2 C33971	ricketsial common
6	276	33.2	159	2 A25972	17K antigen precursor
7	127.5	15.3	182	2 D83169	conserved hypohet
8	120.5	14.5	131	2 A13418	17K surface antigen
9	112.5	13.5	155	2 S23787	outer membrane lip
10	111.5	13.4	154	2 B83514	conserved hypohet
11	109	13.1	179	2 AE0644	hypothetical prote
12	109	13.1	232	2 G87629	hypothetical prote
13	109	13.1	257	2 B82837	conserved hypohet
14	107.5	12.9	142	2 AD2696	lipa protein (limpo
15	107.5	12.9	155	2 AF0289	probable lipoprote
16	107	12.9	125	2 D97478	lipa protein (limpo
17	105.5	12.7	155	2 A10693	outer membrane lip
18	104.5	12.6	155	2 C64921	outer membrane lip
19	104.5	12.6	155	2 F90922	probable outer mem
20	104.5	12.6	155	2 B85771	probable outer mem
21	104	12.5	139	2 S58234	lipa protein - Rhl
22	103	12.4	155	2 AG0443	outer membrane lip
23	103	12.4	232	2 AD3350	outer membrane pro
24	101	12.1	155	2 T64130	hypothetical prote
25	100.5	12.1	257	2 F87413	hypothetical prote
26	100	12.0	179	2 C64855	ycyf protein - Esc
27	100	12.0	179	2 D85674	hypothetical prote
28	100	12.0	179	2 H90814	hypothetical prote
29	99.5	12.0	304	2 H83636	hypothetical prote

30	99	11.9	332	2 D87353	hypothetical prote
31	97	11.7	179	2 AC0198	probable exported
32	97	11.7	223	2 C82230	probable lipoprote
33	95	11.4	172	2 G91049	probable outer mem
34	95	11.4	172	2 H65026	hypothetical prote
35	95	11.4	172	2 D85894	probable outer mem
36	94	11.3	278	2 AB3091	hypothetical prote
37	94	11.3	278	2 H98195	hypothetical prote
38	94	11.3	608	2 T03442	glycine-rich prote
39	92.5	11.1	220	2 G97685	probable outer mem
40	92.5	11.1	220	2 A12910	porin (imported) -
41	92	11.1	100	2 G97672	hypothetical prote
42	92	11.1	100	2 AE2897	conserved hypohet
43	90.5	10.9	691	2 F91251	probable tape meas
44	90	10.8	137	2 B82998	hypothetical prote
45	90	10.8	526	1 KRBOVI	keratin, 54k type

ALIGNMENTS

```
RESULT 1
D33971
ricketsial common antigen precursor - Rickettsia prowazekii
N:Alternate names: 17K surface antigen; outer membrane protein (omp); RP833
C:Species: Rickettsia prowazekii
C>Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 03-Nov-2000
C:Accession: D33971; B71645
R:Anderson, B.E.; Tzianabos, T.
J. Bacteriol. 171, 5199-5201, 1989
A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.
A:Reference number: A33971; M0ID:89359171
A:Accession: D33971
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-159 <AND>
A:Cross-references: GB:M28482; NID:q152461
A:Note: the sequence in GenBank entry R1RANT17KC, release 109.0, (PID:q152462) omits
R:Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; M0ID:99039499
A:Accession: B71645
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-159 <AN2>
A:Cross-references: GB:AJ235273; GB:AJ235269; NID:g3861237; PIDN:CA15258.1; PID:g386
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: omp; RP833
C:Superfamily: rickettsial common antigen
C:Keywords: surface antigen

Query Match          34.6%; Score 288; DB 2; Length 159;
Best Local Similarity 37.9%; Pred. No. 1.7e-18;
Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;

QY 8 SLLIISV---FLVGC--AQNFSRQEVGAATGVGVAGQALFGKSGRVSMAIG-GAVL 61
DB 5 SKMIIITLASMILQACGSGGMKKGITLLGGAGGALLGSGQGQKGL-VGVGVALL 63
QY 62 GGLIGSKIGSMQODK---IKLNSLEKVKAGQVTRMPNPTGNSYSEPPRTYQRYN 117
DB 64 GAVLGQIGASMBEDPRRLLETLSORLLESAPSNSNTEHNPNGNGYVTPNKTY---- 119
QY 118 KOERROGYCREFOQKAMIAQOKOIEYGTACPOPDGRMOVTS 158
DB 120 -RNSAGQYCREYQTVIIGSKQKRTGNACRQPDGGMQVNV 159

RESULT 2
B33971
ricketsial common antigen precursor - Rickettsia conorii
```

C:Species: Rickettsia conorii
 C:Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 20-Aug-1999
 C:Accession: B33971
 R:Anderson, B.E.; Tzianabos, T.
 J. Bacteriol. 171, 5199-5201, 1989
 A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.
 A:Reference number: A33971; MUID:89359171
 A:Accession: B33971
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-159 <AND>
 A:Cross-references: GB:M28479; NID:9152463; PID:AAA26379.1; PID:9152464
 C:Superfamily: rickettsial common antigen

Query Match 34.3%; Score 285; DB 2; Length 159;
 Best Local Similarity 37.9%; Pred. No. 3.1e-18;
 Matches 61; Conservative 31; Mismatches 53; Indels 16; Gaps 6;
 QY 8 SLLIIISV---FLVGC---AQNFSROEVGAATGAVVGAGVAGQLFGKSGRVSMAIG-GAVL 61
 Db 5 SKIMIIATATMLACNGPFGMNKOGTGLTGAGGALLGSGFGKKGQL-VGVGVALL 63
 QY 62 GGLIGSKIGQSMDOQDK----IKLNSLEKVKAGQVTRRNPDGTGNSYSVEPVRTYQRYN 117
 Db 64 GAVLGQIGAGMDEODRRLAELTSGRALETAPSGSNVEMRNPDNGNYGYVTPNKTY----- 119
 QY 118 KOERROQYCREFOOKAMINAGOKOETGTACPOPDGRMOVIS 158
 Db 120 -RNSTGQYCREYTOTYVIGGKQOKAYGNACROPDGMQVYN 159

RESULT 3
 A33971
 Rickettsial common antigen precursor - Rickettsia rickettsii
 C:Species: Rickettsia rickettsii
 C:Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 20-Aug-1999
 C:Accession: A33971
 R:Anderson, B.E.; Tzianabos, T.
 J. Bacteriol. 171, 5199-5201, 1989
 A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.
 A:Reference number: A33971; MUID:89359171
 A:Accession: A33971
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-159 <AND>
 A:Cross-references: GB:M28480; NID:9152457; PID:AAA26376.1; PID:9152458
 C:Superfamily: rickettsial common antigen

Query Match 34.3%; Score 285; DB 2; Length 159;
 Best Local Similarity 37.9%; Pred. No. 3.1e-18;
 Matches 61; Conservative 31; Mismatches 53; Indels 16; Gaps 6;
 QY 8 SLLIIISV---FLVGC---AQNFSROEVGAATGAVVGAGVAGQLFGKSGRVSMAIG-GAVL 61
 Db 5 SKIMIIATATMLACNGPFGMNKOGTGLTGAGGALLGSGFGKKGQL-VGVGVALL 63
 QY 62 GGLIGSKIGQSMDOQDK----IKLNSLEKVKAGQVTRRNPDGTGNSYSVEPVRTYQRYN 117
 Db 64 GAVLGQIGAGMDEODRRLAELTSGRALETAPSGSNVEMRNPDNGNYGYVTPNKTY----- 119
 QY 118 KOERROQYCREFOOKAMINAGOKOETGTACPOPDGRMOVIS 158
 Db 120 -RNSTGQYCREYTOTYVIGGKQOKAYGNACROPDGMQVYN 159

RESULT 4
 G97860
 17K surface antigen precursor [imported] - Rickettsia conorii (strain Malish 7)
 C:Species: Rickettsia conorii
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
 C:Accession: G97860
 R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
 Science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
 A:Reference number: A97700; MUID:21442074; PMID:11557893
 A:Accession: G97860
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-159 <KUR>
 A:Cross-references: GB:AEO06914; PID:AA03825.1; PID:915620425; GSPDB:GN00173
 C:denetics:
 A:gene: omp
 C:Superfamily: rickettsial common antigen

Query Match 34.3%; Score 285; DB 2; Length 159;
 Best Local Similarity 37.9%; Pred. No. 3.1e-18;
 Matches 61; Conservative 31; Mismatches 53; Indels 16; Gaps 6;
 QY 8 SLLIIISV---FLVGC---AQNFSROEVGAATGAVVGAGVAGQLFGKSGRVSMAIG-GAVL 61
 Db 5 SKIMIIATATMLACNGPFGMNKOGTGLTGAGGALLGSGFGKKGQL-VGVGVALL 63
 QY 62 GGLIGSKIGQSMDOQDK----IKLNSLEKVKAGQVTRRNPDGTGNSYSVEPVRTYQRYN 117
 Db 64 GAVLGQIGAGMDEODRRLAELTSGRALETAPSGSNVEMRNPDNGNYGYVTPNKTY----- 119
 QY 118 KOERROQYCREFOOKAMINAGOKOETGTACPOPDGRMOVIS 158
 Db 120 -RNSTGQYCREYTOTYVIGGKQOKAYGNACROPDGMQVYN 159

RESULT 5
 C33971
 Rickettsial common antigen precursor - Rickettsia typhi
 C:Species: Rickettsia typhi
 C:Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 20-Aug-1999
 C:Accession: C33971
 R:Anderson, B.E.; Tzianabos, T.
 J. Bacteriol. 171, 5199-5201, 1989
 A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.
 A:Reference number: A33971; MUID:89359171
 A:Accession: C33971
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-159 <AND>
 A:Cross-references: GB:M28481; NID:9152459; PID:AAA26377.1; PID:9152460
 C:Superfamily: rickettsial common antigen

Query Match 33.2%; Score 276.5; DB 2; Length 159;
 Best Local Similarity 39.6%; Pred. No. 1.8e-17;
 Matches 55; Conservative 27; Mismatches 46; Indels 11; Gaps 4;
 QY 25 SROEVGAATGAVVGAGVAGQLFGKSGRVSMAIG-GAVLGGLIGSKIGQSMDOQDK---I 79
 Db 27 NKOGTGLTGAGGALLGSGFGKKGQL-VGVGVALGAVLGQIGAGLDEODRKLLEL 85
 QY 80 KLNQSLKVKAGQVTRRNPDGTGNSYSVEPVRTYQRYNKOERROQYCREFOOKAMINAGOK 139
 Db 86 TSGRALESAPSGSNVEMRNPDNGNYGYVTPNKTY-----RNSTGQYCREYTOTYVIGGKQ 140
 QY 140 QETGTACPOPDGRMOVIS 158
 Db 141 QETYNACROPDGMQVYN 159

RESULT 6
 A25972
 17K antigen precursor - Rickettsia rickettsii
 C:Species: Rickettsia rickettsii
 C:Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 20-Aug-1999
 C:Accession: A25972
 R:Anderson, B.E.; Regnery, R.L.; Carlone, G.M.; Tzianabos, T.; McDade, J.E.; Fu, Z.Y.
 J. Bacteriol. 169, 2385-2390, 1987
 A:Title: Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia ricketts
 A:Reference number: A25972; MUID:87222152
 A:Accession: A25972

A:Molecule type: DNA
A:Residues: 1-159 <AND>
A:Cross-references: GB:M16486; NID:g152467; PIDN:AAA6381.1; PID:g152468
C:superfamily: rickettsial common antigen

Query Match 33.2%; Score 276; DB 2; Length 159;
Best Local Similarity 37.3%; Pred. No. 2e-17;
Matches 60; Conservative 31; Mismatches 54; Indels 16; Gaps 6;
QY 8 SLLIIISV---FLVGC--AONFSROEVGAATGAVGVAGQLFGKSGRYSMAIG-GAVL 61
D 5 SKLIIIALATSMQLQACNPGCMNKGCTGLLGAGAGALLGSGFGKGGQL-VGVGVGALL 63
QY 62 GGLIGSGISGSMDOOK-----IKLNSLEKVRKAGQVTRMRNPPTGNSYSVEPYRTYQRVN 117
D 64 GAVLGQIGAGHDEQDRRLAELTSGRALETAPEGSGSNVEMRNPNNGNGYTPPKTY----- 119
QY 118 KQERRQOYCFEFGQKAMIAQKQEIYGTACPPDPGRMNVIS 158
D 120 -RNSTGOYCREYQTQTVVIGGKQKAYGACRQPDQEDQOVVN 159

RESULT 7
D83169
conserved hypothetical protein PA3819 [Imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83169
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lartig, K.; Lam,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MID:20437337
A:Accession: D83169
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-182 <STO>
A:Cross-references: GB:AE004799; GB:AE004091; NID:g9949981; PIDN:AA607206.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3819

Query Match 15.3%; Score 127.5; DB 2; Length 182;
Best Local Similarity 34.8%; Pred. No. 0.00037;
Matches 32; Conservative 13; Mismatches 28; Indels 19; Gaps 2;
QY 27 QEVGAATGAVGVAGQLFGKSGRYSMAIGAVLGGLIGSKIGSMDQDKI----- 79
D 70 QINGTAIGAVVGGLLNQGCGTGKRIATYAGAVGGGYNKQVEGMDERDVTITETRC 129
QY 80 -KLNSLEK-----KAGQVTRMRNP 99
D 130 STVHDSSEKVVGYDVKYMLDGKAGQIRMERDP 161

RESULT 8
A13418
17K surface antigen precursor [Imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: A13418
R:Delvecchio, V.G.; Kaparatel, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: AD3252; PMID:11756688
A:Accession: A13418
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-131 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL52516.1; PID:g17983328; GSPDB:GN00190
A:Experimental source: strain 16M

C:Genetics:
A:Gene: BME11335
A:Map position: I

Query Match 14.5%; Score 120.5; DB 2; Length 131;
Best Local Similarity 31.7%; Pred. No. 0.0011;
Matches 39; Conservative 17; Mismatches 44; Indels 23; Gaps 7;
QY 46 GKSSGRYSMAIG-----AVLG-GLIGSKIGQ--SMDQDKIKLN-QSLEKVKAG 91
D 14 GKSSGFPSS--LGSSSQKPEPTNLLASLNGLFGNSASQSLAADRKKLAELRYALVSPAG 71
QY 92 QVTRMRNPDTGNSYSVEPYRTYQRVKNQERRQOYCFEFGQKAMIAQKQEIYGTACPPDP 151
D 72 KSVLMSGAGS-NAGDVTAAQPIQ-----VGSQNCROYSHSFTITGGDOQIVRGTAACRNP 124
QY 152 GRW 154
D 125 GSW 127

RESULT 9
S23787
outer membrane lipoprotein precursor - Yersinia enterocolitica
C:Species: Yersinia enterocolitica
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 29-Sep-1999
C:Accession: S23787
R:Baeumler, A.J.; Hantke, K.
J. Bacteriol. 174, 1029-1035, 1992
A:Title: A lipoprotein of Yersinia enterocolitica facilitates ferrioxamine uptake in
A:Reference number: S23786; MID:92121089
A:Accession: S23787
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-155 <BAE>
A:Cross-references: EMBL:X60448; NID:g48577; PIDN:CAA42977.1; PID:g48579
C:Superfamily: PAL cross-reacting lipoprotein

Query Match 13.5%; Score 112.5; DB 2; Length 155;
Best Local Similarity 24.7%; Pred. No. 0.0068;
Matches 37; Conservative 19; Mismatches 43; Indels 51; Gaps 4;
QY 9 SLLIIISVFLVGCAN-----FSROE-----VGAAT 33
D 7 AVAIIAAVTLTGCAANNNTLSGDVFSASQAKOYQVTTGTLISVPVITIGGDDNNWGAIG 66
QY 34 GAVVGVAGQLFGKSGRYSMAIGAVLGGLIGSKIGSMDQDKIKL----- 81
D 67 GAVLGFLGNTVGGTGRSLATAAGAVAGGACGAGGANNRTDGVQLEVRKDDGTTILV 126
QY 82 --NQSLEKAKQVTRMRNPDTGNSYSVEP 109
D 127 VQKQGPTRFSVQ--RVMLASGSGTIVTSP 154

RESULT 10
B83514
conserved hypothetical protein PA1053 [Imported] - Pseudomonas aeruginosa (strain PAO
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83514
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
radman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lartig, K.; L
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MID:20437337
A:Accession: B83514
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <STO>
A:Cross-references: GB:AE004537; GB:AE004091; NID:g99446960; PIDN:AA604442.1; GSPDB:GN
A:Experimental source: strain PA01

Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tso, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AD2696
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-142 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AAL41986.1; PID:q17739356; GSPDB:GM00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: lipA
 A:Map position: circular chromosome

THIS PAGE BLANK (USPTO)

Db 491 GGAGAAAAAAGAGGEGIRGGAGGGYGLGSGSGRGLGGGAGAAAAAAGAGGGG 550
QY 64 LIGSKIGQ 71
Db 551 LGGGAGGQ 558

RESULT 2
US-09-247-806-1

Sequence 1, Application US/09247806
Patent No. 6280747
GENERAL INFORMATION:
APPLICANT: PHILIPPE, Michel
APPLICANT: GARSON, Jean-Claude
APPLICANT: ARRAUDEAU, Jean-Pierre
TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
FILE REFERENCE: 6388-0365-0
CURRENT FILING DATE: 1999-02-11
EARLIER APPLICATION NUMBER: US/09/247,806
EARLIER FILING DATE: 1998-02-11
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 1
LENGTH: 651
TYPE: PRT
ORGANISM: Nephila clavipes
US-09-247-806-1

Query Match
Best Local Similarity 9.4%; Score 78.5; DB 4; Length 651;
Matches 24; Conservative 4; Mismatches 17; Indels 23; Gaps 2;

QY 27 QEVGATGAVVG---GVAGQ-----LFGKSGRVSMAIGAVLGG 63
Db 491 GGAGAAAAAAGAGGEGIRGGAGGGYGLGSGSGRGLGGGAGAAAAAAGAGGGG 550
QY 64 LIGSKIGQ 71
Db 551 LGGGAGGQ 558

RESULT 3

US-08-425-069-2
Sequence 2, Application US/08425069
Patent No. 5728810
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
APPLICANT: Himman, Michael B.
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5728810th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,069
FILING DATE: 19-APR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-069-2

Query Match
Best Local Similarity 9.4%; Score 78.5; DB 1; Length 718;
Matches 24; Conservative 4; Mismatches 17; Indels 23; Gaps 2;

QY 27 QEVGATGAVVG---GVAGQ-----LFGKSGRVSMAIGAVLGG 63
Db 491 GGAGAAAAAAGAGGEGIRGGAGGGYGLGSGSGRGLGGGAGAAAAAAGAGGGG 550
QY 64 LIGSKIGQ 71
Db 551 LGGGAGGQ 558

RESULT 4

US-08-317-844B-2
Sequence 2, Application US/08317844B
Patent No. 5989894
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
APPLICANT: Himman, Michael B.
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5989894th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,844B
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 241-1300
TELEFAX: (703) 241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-317-844B-2

Query Match 9.4%; Score 78.5; DB 2; Length 718;

Best Local Similarity 35.3%; Pred. No. 1.9;
Matches 24; Conservative 4; Mismatches 17; Indels 23; Gaps 2;

OY 27 QEVGAATGAVG-----GVAGQ-----LFGKSGSRVSMATGAVLGG 63
Db 491 QGAGAAAAAAGAGGEGIRGAGAGGGYGLGSGSGRGLGGGAGAAAAAGAGAGGG 550

OY 64 LIGSKIGQ 71
Db 551 LGGGAGAG 558

RESULT 5

US-09-034-177-3
; Sequence 3, Application US/09034177

; Patent No. 6127146

; GENERAL INFORMATION:

; APPLICANT: Lal, Preeti

; APPLICANT: Guegler, Karl J.

; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: HUMAN FIBROUS PROTEIN

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/034,177

; FILING DATE: HEREWITH

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0486 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 747 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLOVE: GI 1174414

US-09-034-177-3

Query Match 9.4%; Score 78.5; DB 3; Length 747;

Best Local Similarity 35.3%; Pred. No. 2;

Matches 24; Conservative 4; Mismatches 17; Indels 23; Gaps 2;

OY 27 QEVGAATGAVG-----GVAGQ-----LFGKSGSRVSMATGAVLGG 63
Db 491 QGAGAAAAAAGAGGEGIRGAGAGGGYGLGSGSGRGLGGGAGAAAAAGAGAGGG 550

OY 64 LIGSKIGQ 71
Db 551 LGGGAGAG 558

RESULT 6
US-09-553-498-8
; Sequence 8, Application US/09553498

; Patent No. 6309861

; GENERAL INFORMATION:

; APPLICANT: Ambrosius, Dorothee

; APPLICANT: Rudolph, Rainer

; APPLICANT: Schaeffner, Joerg

; APPLICANT: Schwarz, Elisabeth

; TITLE OF INVENTION: Process for the production of naturally folded and secreted pr

; FILE REFERENCE: Case 20379

; CURRENT APPLICATION NUMBER: US/09/553,498

; CURRENT FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: EP99107412.1

; PRIOR FILING DATE: 1999-04-26

; NUMBER OF SEQ ID NOS: 10

; SEQ ID NO 8

; LENGTH: 255

; TYPE: PR

; ORGANISM: E. coli

US-09-553-498-8

Query Match 9.3%; Score 77.5; DB 4; Length 255;

Best Local Similarity 20.9%; Pred. No. 0.62;

Matches 37; Conservative 28; Mismatches 81; Indels 31; Gaps 7;

OY 5 LGGSLIITISVFLVGCANFSRQEVGAATGAV-----GVAGQLFGKSGSRVSMATGAV 60

Db 83 LQMTSLRSEDYATVYCCARDYGAY-WGGGTVTYVSSGGGSGGSGGSDIELTSPAI 141

OY 61 LGGIGSKIGQSMQDQKIR-LN-----QSLKVKAGGVTRMRNDTGSY 105

Db 142 MSASPGKVTMTCSASSVRYMMNFQOKSGTSPKRMITDTSKLSGVPARFSGSGSTSY 201

OY 106 SVBPRTYORYNKOERQYCRFQOKAMI--AGQKQIYGTACPPDGRMNYISTE 160

Db 202 SL-----TISMEADATYTCQOWSNPLTFEGATKLEKRAAEQ-----KLISSE 249

RESULT 7

US-09-067-351-2

; Sequence 2, Application US/09067351

; Patent No. 5994081

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Corley, Neil C.

; APPLICANT: Baughn, Mariah

; TITLE OF INVENTION: HUMAN KERATINS

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 PORTER DRIVE

; CITY: PALO ALTO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/067,351

; FILING DATE: Herewith

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: CERRONE, MICHAEL C.

; REGISTRATION NUMBER: 39,132

; REFERENCE/DOCKET NUMBER: PF-0511 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (650) 855-0555

; TELEFAX: (650) 845-4166

DB 2448 SPIIPDNRRLRRVATVTTTNNNNKSOVSQNNSSLLVNRANANSONMNSPTGQVQOOSPLR 2507
QY 137 GOKOEIYGT 145
DB 2508 GOGNQTYS 2516

RESULT 10

US-08-895-590-2
; Sequence 2, Application US/08895590
; Patent No. 6207410
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; APPLICANT: Ren, Dejian
; APPLICANT: Zheng, Wei
; APPLICANT: Dubaig, Manuel Marcel Paul
; TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,590
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,888
; FILING DATE: 19-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm M.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 022650-263
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2516 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-895-590-2

Query Match 9.2%; Score 76.5; DB 4; Length 2516;
Best Local Similarity 24.8%; Pred. No. 17;
Matches 32; Conservative 18; Mismatches 66; Indels 13; Gaps 4;

QY 29 VGAAATGAVVGVAGQLFGKSGRYMAIGG-AVLGLISKITGQS-----MDQODKIKLN 82
DB 2389 IGSSNSGIFGSGAGCGAGCGVG-GLGSSSIRNAFGSGSGSPSLSPQHOPYSGLTN 2447
QY 83 -----QSLEKVKAGQVTRMRNDPTGNSVSEVRYQRYNKQERROQYREFQOKMIA 136
DB 2448 SPIIPDNRRLRRVATVTTTNNNNKSOVSQNNSSLLVNRANANSONMNSPTGQVQOOSPLR 2507
QY 137 GOKOEIYGT 145
DB 2508 GOGNQTYS 2516

RESULT 11
US-07-803-633A-13
; Sequence 13, Application US/07803633A
; Patent No. 5369025

; GENERAL INFORMATION:
; APPLICANT: NAZERIAN, Keyvan
; APPLICANT: LEE, Lucy F.
; APPLICANT: YANAGIDA, No. 53690250ru
; APPLICANT: OGAMA, Ryohel
; APPLICANT: LI, YI
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VACCINE FOR
; PROTECTION AGAINST MAREK'S DISEASE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 301 No. 5369025th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/803,633A
; FILING DATE: 19911210
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1644-103P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 865 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-803-633A-13

Query Match 9.1%; Score 76; DB 1; Length 865;
Best Local Similarity 25.9%; Pred. No. 4.7;
Matches 37; Conservative 30; Mismatches 46; Indels 30; Gaps 8;

QY 16 FLVCAQNESRQ-EVGAATGAVVGVAGQLFGKSGRVS-----MAIGAVLGLI 65
DB 688 FMNGLAEFLFMGQVGAIGKVVVGAAGAIVSTISG-VSAFMSNPGALIGLIIAGLV 746
QY 66 GSKIGSQMDQODKIKLN-----QSLEKVKAGQVTRMRNDPTGNSVSEVRYQRYNK 118
DB 747 AAFVLAFL-AYRYVNRKLSPMKALYPMTEVLKA-QATRELHESDD-----LERTSI 795
QY 119 QERRQCYRE-FOOKAMIAQOKQ 140
DB 796 DERKLEAREMITYMALVSAER 818

RESULT 12
US-08-220-151-6
; Sequence 6, Application US/08220151
; Patent No. 5529780
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Limbach, Keith J.
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
; CANINE HERPESVIRUS GB, GC AND GD AND USES THEREFOR
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY

COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/220,151
FILING DATE: 30-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FROMMER, WILLIAM S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2540
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
US-08-220-151-6

Query Match 9.0%; Score 74.5; DB 1; Length 913;
Best Local Similarity 23.6%; Pred. No. 7.4;
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

QY 15 VFLVCAQNP-SRQEVGAATGAVGVAGQLFGKSGRVS-----MAIGAVLGGLI 65
DB 753 VLLRGIANFQGLGDVGAAGVKVILGATGAVISAVGAWSFSLNFGALALGLVLGLV 812
QY 66 GSKT-----GQSMDDQDKIKLNSLEKVR-----A 90
DB 813 AAFLAYRHISRLRNPMKALYPVTTKLEKEDVDEGDVDEAKLDQARDMIRYMSIVSALE 872
QY 91 GQVTRRNPDPTGNSYSVEPV-----RTYQRYNKOE 120
DB 873 QOEHRKRRKNSGPAALLASRVGAMATRRRHRYQLSESD 909

RESULT 13
US-08-413-118-6
Sequence 6, Application US/08413118
Patent No. 5688920
GENERAL INFORMATION:
APPLICANT: PAOLETTI, ENZO
APPLICANT: LIMBACH, KEITH J.
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
NUMBER OF INVENTION: CANINE HERPESVIRUS 9B, 9C, AND 9D AND USES THEREFOR
NUMBER OF SEQUENCES: 128
CORRESPONDENCE ADDRESS:
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.
STREET: 530 FIFTH AVENUE, 25TH FLOOR
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/413,118
FILING DATE: 29-MAR-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/220,151
FILING DATE: 30-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: FROMMER, WILLIAM S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2670
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
US-08-413-118-6

Query Match 9.0%; Score 74.5; DB 1; Length 913;
Best Local Similarity 23.6%; Pred. No. 7.4;
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

QY 15 VFLVCAQNP-SRQEVGAATGAVGVAGQLFGKSGRVS-----MAIGAVLGGLI 65
DB 753 VLLRGIANFQGLGDVGAAGVKVILGATGAVISAVGAWSFSLNFGALALGLVLGLV 812
QY 66 GSKT-----GQSMDDQDKIKLNSLEKVR-----A 90
DB 813 AAFLAYRHISRLRNPMKALYPVTTKLEKEDVDEGDVDEAKLDQARDMIRYMSIVSALE 872
QY 91 GQVTRRNPDPTGNSYSVEPV-----RTYQRYNKOE 120
DB 873 QOEHRKRRKNSGPAALLASRVGAMATRRRHRYQLSESD 909

RESULT 14
US-08-473-446-6
Sequence 6, Application US/08473446
Patent No. 6017542
GENERAL INFORMATION:
APPLICANT: PAOLETTI, ENZO
APPLICANT: LIMBACH, KEITH J.
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
NUMBER OF INVENTION: CANINE HERPESVIRUS 9B, 9C, AND 9D AND USES THEREFOR
NUMBER OF SEQUENCES: 128
CORRESPONDENCE ADDRESS:
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.
STREET: 530 FIFTH AVENUE, 25TH FLOOR
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,446
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/413,118
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: FROMMER, WILLIAM S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2670
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
US-08-473-446-6

Query Match 9.0%; Score 74.5; DB 3; Length 913;
Best Local Similarity 23.6%; Pred. No. 7.4;
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

OY 15 VFLVGCANF-SROEVGAATGAVVGAVGOLFSGSGRVS-----MAIGAVLGGLI 65
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 753 VLLRGIANFPGLDVGAAGKVVVGATGAVISAVGKWSFLSNPFGALATGLVLAGLV 812
OY 66 GSKI-----GQSMDDQDKIKLNSLEKVK-----A 90
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 813 AAFLAYRHISRLRRNPKKALYPVTTKLEKGVDEGVDEAKLDQARDMIRMSIVSALE 872
OY 91 GQVTRWRNPDTGNSYSVEPV-----RTYQRYNKOE 120
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 873 QOEHKARKKNSGALLASRVGAMATRRRHVQRLSESD 909

RESULT 15
US-09-232-468A-2
; Sequence 2, Application US/09232468A
; Patent No. 6207165
; GENERAL INFORMATION:
; APPLICANT: AUDONNET et al.
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA AGAINST PORCINE
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY PATHOLOGIES
; FILE REFERENCE: 454313-2230
; CURRENT APPLICATION NUMBER: US/09/732.468A
; CURRENT FILING DATE: 1999-01-05
; NUMBER OF SEQ. ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 913
; TYPE: PRT
; ORGANISM: Pseudorabies virus
US-09-232-468A-2

Query Match 9.0%; Score 74.5; DB 4; Length 913;
Best Local Similarity 23.6%; Pred. No. 7.4;
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

OY 15 VFLVGCANF-SROEVGAATGAVVGAVGOLFSGSGRVS-----MAIGAVLGGLI 65
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 753 VLLRGIANFPGLDVGAAGKVVVGATGAVISAVGKWSFLSNPFGALATGLVLAGLV 812
OY 66 GSKI-----GQSMDDQDKIKLNSLEKVK-----A 90
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 813 AAFLAYRHISRLRRNPKKALYPVTTKLEKGVDEGVDEAKLDQARDMIRMSIVSALE 872
OY 91 GQVTRWRNPDTGNSYSVEPV-----RTYQRYNKOE 120
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 873 QOEHKARKKNSGALLASRVGAMATRRRHVQRLSESD 909

Search completed: October 27, 2002, 11:12:13
Job time : 9.34197 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:58:51 : Search time 21.4111 Seconds
(without alignments)
835.217 Million cell updates/sec

Title: US-09-677-374-4
Perfect score: 832
Sequence: 1 MRGCLGSSLLIISVFLVGC.....ITGTACPPDGRMQUISTFEK 161

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq-032802.*
1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	832	100.0	161	22	AA881127
2	832	100.0	236	22	AA881128
3	815	98.0	162	22	AA881126
4	815	98.0	162	22	AA881126
5	112	13.5	20	22	AA881130
6	111	13.3	224	22	AA881130
7	102.5	12.3	223	20	AA134487
8	102.5	12.3	230	20	AA134487
9	101	12.1	154	11	AA805799
10	95	11.4	309	22	AA815906
11	88	10.6	528	22	AA882611

12	87.5	10.5	147	22	AB869847
13	86.5	10.4	2309	22	AB866232
14	84	10.1	1251	22	AB861254
15	83	10.0	116	19	AA111028
16	83	10.0	423	22	AB830695
17	82	9.9	2017	22	AB806301
18	81	9.7	666	22	AB858019
19	80.5	9.7	542	22	AB857900
20	80.5	9.7	542	22	AB857901
21	80.5	9.7	542	22	AB870501
22	79.5	9.6	618	21	AB856803
23	78.5	9.4	651	20	AA140097
24	78.5	9.4	718	12	AA14308
25	78.5	9.4	718	19	AA853346
26	78.5	9.4	718	21	AA59070
27	77.5	9.3	102	22	AA41943
28	77.5	9.3	255	21	AA811398
29	77.5	9.3	255	22	AA874199
30	77.5	9.3	255	22	AA870769
31	77.5	9.3	255	22	AA772020
32	77.5	9.3	285	20	AA41688
33	77.5	9.3	285	21	AB844244
34	77.5	9.3	285	22	AA029025
35	77.5	9.3	285	22	AA839011
36	77.5	9.3	302	22	AA40157
37	77.5	9.3	354	22	AA19445
38	77.5	9.3	514	22	AA36520
39	77	9.3	223	21	AA857121
40	77	9.3	285	21	AA873440
41	77	9.3	285	21	AA873440
42	77	9.3	285	22	AA693288
43	77	9.3	551	21	AA52398
44	77	9.3	2599	21	AA75098
45	76.5	9.2	2516	17	AAW01875

ALIGNMENTS

RESULT 1	
AA881127	
ID	AA881127 standard; Protein: 161 AA.
XX	
XX	AA881127;
XX	
DT	11-JUL-2001 (first entry)
XX	
DE	Optimised Ospa protein 17E2 amino acid sequence.
KW	Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW	vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
KW	SRS.
XX	
OS	Piscirickettsia salmonis.
OS	Synthetic.
XX	
FT	Key
FT	Region
XX	
PN	CA281913-A1.
XX	
PD	17-MAR-2001.
XX	
XX	17-SEP-1999; 99CA-2281913.
PF	
XX	
XX	17-SEP-1999; 99CA-2281913.
PR	
XX	
XX	
PA	(KAYW) KAY W W.
PA	(BURJ) BURIAN J.
PA	(KUZV) KUZV M A.
XX	
PI	Key WW, Burian J, Kuzv MA;

Drosophila melanog
Drosophila melanog
Drosophila melanog
H. pylori ORF 01cp
A fusion of anti-C
Novel human diagno
Drosophila melanog
Drosophila melanog
Drosophila melanog
Drosophila melanog
Human prostate can
Spider silk protei
N. clavipes draglin
Nephila clavipes s
N. clavipes spider
Human polyptide
E. coli expression
Peib-scfoxaolon
Expression plasmid
E. carotovora Peib
Human PRO284 (UNO2
Human PRO polypt
Human polyptide
Human polyptide
Pseudomonas aerugi
Human prostate can
Human signal pepti
Human secreted pro
Human protein HP10
Human keratin KERT
Neisseria meningit
Neuronal Invertebr

XX WPI: 2001-316844/34.
DR N-PSDB: AAF86247.

XX Method for protecting polkilohermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
XX vaccine containing the Ospa protein of *Piscirickettsia salmonis*.
XX
XX Disclosure: Fig 5; 35pp; English.

CC This invention relates to a method for the protection against infection
CC of a polkilohermic fish by the bacterial pathogen, *Piscirickettsia*
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkilohermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents optimised P. salmonis Ospa protein 17E2. The DNA
CC encoding Ospa 17E2 (AAF86247) has been optimised for expression in
CC *Escherichia coli*. An Ospa protein with an N-terminal fusion partner is
CC used in a vaccine to create an anti-Ospa antibody response.

SQ Sequence 161 AA:

Query Match 100.0%; Score 832; DB 22; Length 161;
Best Local Similarity 100.0%; Pred. No. 3e-81;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MRCCLOGSSLIITISVFLVGCACNFRSEYGAATGAVGVAGQLFGKSGRYMAIGAV 60
DB 1 MRCCLOGSSLIITISVFLVGCACNFRSEYGAATGAVGVAGQLFGKSGRYMAIGAV 60

OY 61 LGGTIGSKTIGQSMDDQDKIKLNQSLKRYKAGVTRMRNPDGNSYSVPVRYRYNKOE 120
DB 61 LGGTIGSKTIGQSMDDQDKIKLNQSLKRYKAGVTRMRNPDGNSYSVPVRYRYNKOE 120

OY 121 RROQYCREFOQKAMTAGOKOEITGTACPOPDRMOWISTEK 161
DB 121 RROQYCREFOQKAMTAGOKOEITGTACPOPDRMOWISTEK 161

RESULT 2
AAB81128
ID AAB81128 standard; Protein: 256 AA.

AC AAB81128;
DT 11-JUL-2001 (first entry)

DE C17E2 Ospa construct with N-terminal fusion partner.

XX Polkilohermic fish; *Piscirickettsia salmonis*; rickettsial pathogen;
KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
XX SRS; 17E2; fusion construct.

OS *Piscirickettsia salmonis*.
XX Synthetic.

XX Key location/Qualifiers
FH 1..95
FT /label= Undefined_N-terminal_fusion_partner
FT /label= C17E2_Ospa
FT /label= "Product of Ospa gene optimised for expression in
FT /note= "Product of Ospa gene optimised for expression in
XX *Escherichia coli*"

CA2281913-A1.

17-MAR-2001.

17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

XX (KAYW/) KAY W W.
PA (BURJ/) BURIAN J.
PA (KUZV/) KUZYK M A.

XX Kay WW, Burian J, Kuzyk MA;

DR WPI: 2001-316844/34.
DR N-PSDB: AAF86248.

XX Method for protecting polkilohermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
XX vaccine containing the Ospa protein of *Piscirickettsia salmonis*.
XX
XX Example 4; Fig 5; 35pp; English.

CC This invention relates to a method for the protection against infection
CC of a polkilohermic fish by the bacterial pathogen, *Piscirickettsia*
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkilohermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents the amino acid sequence of C17E2 a P. salmonis Ospa
CC construct optimised for expression in *Escherichia coli*, fused to an
CC undefined N-terminal fusion partner. The fusion protein is used in a
CC vaccine to create an anti-Ospa antibody response.

SQ Sequence 256 AA:

Query Match 100.0%; Score 832; DB 22; Length 256;
Best Local Similarity 100.0%; Pred. No. 5.4e-81;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MRCCLOGSSLIITISVFLVGCACNFRSEYGAATGAVGVAGQLFGKSGRYMAIGAV 60
DB 96 MRCCLOGSSLIITISVFLVGCACNFRSEYGAATGAVGVAGQLFGKSGRYMAIGAV 155

OY 61 LGGTIGSKTIGQSMDDQDKIKLNQSLKRYKAGVTRMRNPDGNSYSVPVRYRYNKOE 120
DB 156 LGGTIGSKTIGQSMDDQDKIKLNQSLKRYKAGVTRMRNPDGNSYSVPVRYRYNKOE 215

OY 121 RROQYCREFOQKAMTAGOKOEITGTACPOPDRMOWISTEK 161
DB 216 RROQYCREFOQKAMTAGOKOEITGTACPOPDRMOWISTEK 256

RESULT 3
AAG78025

ID AAG78025 standard; Protein: 162 AA.

AC AAG78025;

DT 15-JAN-2002 (first entry)

DE *Piscirickettsia salmonis* polypeptide P10.6.

XX *Piscirickettsia salmonis*; *Piscirickettsiosis*; salmonid rickettsial;
KW septicaemia; SRS; surface antigen; vaccine; antibacterial; fish;
XX ATCC VR-1361.

XX *Piscirickettsia salmonis*.

OS WO200168865-A2.

20-SEP-2001.

12-MAR-2001; 2001WO-GB01055.

11-MAR-2000; 2000GB-0005838.
PR 01-JUL-2000; 2000GB-0016080.

```

PR 01-JUL-2000; 2000GB-0016082.
PR 29-JUL-2000; 2000GB-0018599.
XX
XX (AQUA-) AQUA HEALTH EURO LTD.
XX
XX Simard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;
PI Burzio L;
XX WPI: 2001-639050/73.
DR N-PSDB; AAF79040.
XX
XX New nucleic acids encoding an amino acid sequence homologous to the
PT surface antigen present on piscirickettsia salmonis are useful to
PT protect fish against piscirickettsiosis -
XX
XX Claim 6; Fig 5; 25pp; English.
XX
XX The invention relates to nucleic acid sequences and the encoded protein
CC of a least part of the surface antigen present on piscirickettsia
CC salmonis for production of a vaccine with antibacterial activity to
CC protect fish against P. salmonis which causes piscirickettsiosis, also
CC known as salmonid rickettsial septicaemia.
XX
SQ Sequence 162 AA:

Query Match          98.0%; Score 815; DB 22; Length 162;
Best Local Similarity 98.8%; Pred. No. 2e-79;
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 RGLQSSLLIIIVFLVGCQNFSROEVGAATGAVVGGVAGOLFQKSGSVMAIGAVL 61
DB 3 RGLQSSLLIIIVFLVGCQNFSROEVGAATGAVVGGVAGOLFQKSGSVMAIGAVL 62
OY 62 GGLIGSKIGSQMDQDKIKLNQSLKLVKAGQVTRMRNPDTGNSYSVEPVRTYGRYNKQER 121
DB 63 GGLIGSKIGSQMDQDKIKLNQSLKLVKAGQVTRMRNPDTGNSYSVEPVRTYGRYNKQER 122
OY 122 ROOYCREFOOKAMIAQOKOEIYGTACPDGGRMOVISTEK 161
DB 123 ROOYCREFOOKAMIAQOKOEIYGTACPDGGRMOVISTEK 162

RESULT 4
AAB81126
ID AAB81126 standard; Protein; 162 AA.
XX
XX AAB81126;
AC
XX
XX 11-JUL-2001 (first entry)
DT
XX
XX OsPA antigen amino acid sequence.
DE
XX
XX polkilohermic fish; piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OsPA; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS.
XX
XX Piscirickettsia salmonis.
OS
XX
XX Key Location/Qualifiers
FH 110..129
FT Region /label= B_cell_epitope
XX
XX CA2281913-A1.
PN
XX
XX 17-MAR-2001.
PD
XX
XX 17-SEP-1999; 99CA-2281913.
PF
XX
XX 17-SEP-1999; 99CA-2281913.
PR
XX
XX (KAWW/) KAY W W.
PA (BURJ/) BURIAN J.
PA (KUZV/) KUZYK M A.
XX

```

```

XX
XX Kay WW, Burian J, Kuzyk MA;
PI
XX
XX WPI: 2001-316844/34.
DR N-PSDB; AAF86246.
XX
XX Method for protecting polkilohermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the OsPA protein of piscirickettsia salmonis -
XX
XX Example 2; Fig 2B; 35pp; English.
XX
XX This invention relates to a method for the protection against infection
CC of a polkilohermic fish by the bacterial pathogen, piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed OsPA, or an immunogenic fragment of
CC OsPA in the form of a vaccine. The method is used for protecting animals,
CC particularly polkilohermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents P. salmonis OsPA protein. An OsPA protein with an
CC N-terminal fusion partner is used in a vaccine to create an anti-OsPA
CC antibody response.
XX
SQ Sequence 162 AA:

Query Match          98.0%; Score 815; DB 22; Length 162;
Best Local Similarity 98.8%; Pred. No. 2e-79;
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 RGLQSSLLIIIVFLVGCQNFSROEVGAATGAVVGGVAGOLFQKSGSVMAIGAVL 61
DB 3 RGLQSSLLIIIVFLVGCQNFSROEVGAATGAVVGGVAGOLFQKSGSVMAIGAVL 62
OY 62 GGLIGSKIGSQMDQDKIKLNQSLKLVKAGQVTRMRNPDTGNSYSVEPVRTYGRYNKQER 121
DB 63 GGLIGSKIGSQMDQDKIKLNQSLKLVKAGQVTRMRNPDTGNSYSVEPVRTYGRYNKQER 122
OY 122 ROOYCREFOOKAMIAQOKOEIYGTACPDGGRMOVISTEK 161
DB 123 ROOYCREFOOKAMIAQOKOEIYGTACPDGGRMOVISTEK 162

RESULT 5
AAB81130
ID AAB81130 standard; Peptide; 20 AA.
XX
XX AAB81130;
AC
XX
XX 11-JUL-2001 (first entry)
DT
XX
XX OsPA B-cell epitope peptide #2.
DE
XX
XX polkilohermic fish; piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OsPA; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS; antibody.
XX
XX Piscirickettsia salmonis.
OS
XX
XX Key Location/Qualifiers
FH 110..129
FT Region /label= B_cell_epitope
XX
XX CA2281913-A1.
PN
XX
XX 17-MAR-2001.
PD
XX
XX 17-SEP-1999; 99CA-2281913.
PF
XX
XX 17-SEP-1999; 99CA-2281913.
PR
XX
XX (KAWW/) KAY W W.
PA (BURJ/) BURIAN J.
PA (KUZV/) KUZYK M A.
XX
XX Kay WW, Burian J, Kuzyk MA;
PI
XX

```

Example 2: Page 17: 35pp; English.

This invention relates to a method for the protection against infection of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia salmonis*. The method comprises administering an immunogenic amount of *P. salmonis* specific antigen termed OspA, or an immunogenic fragment of OspA in the form of a vaccine. The method is used for protecting animals, particularly poikilothermic fish, against the bacterial pathogen *P. salmonis*. The method is also useful for protecting against salmonid rickettsial septicemia (SRS) and other rickettsial diseases. The sequence represents an immunogenic epitope of the *P. salmonis* OspA protein. The peptide is used to raise rabbit anti-OspA antibodies.

Sequence 20 AA;

```

Query Match      13.5%; Score 112; DB 22; length 20;
Best Local Similarity 100.0%; Pred. No. 3.3e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 109 PVRTYRYNKKQKRRQOYCRE 128
Db 1 PVRTYRYNKKQKRRQOYCRE 20

```

RESULT 6	
AAB20105	
ID	AAB20105 standard; Protein; 224 AA
XX	

43-APR-2001 (first entry)

calcrinalis BASB113 protein.

antibacterial; antimicrobial.

Moraxella catarrhalis.

04-JAN-2001.

23-JUN-2000; 2000WO-EP05851.

25-JUN-1999; 99GB-0015044.

BECHAM BIOLOGICALS

Thonnard J;

N-PSDB; AAF30043

N-PSDB; AAF30043

Claim 1, Page 67, 86pp, English.

The amino acid sequence is that of BasiB13 protein from Moraxella catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis media in children and pneumonia in adults. The invention provides BasiB13 polypeptides, and polynucleotides encoding them, as well as expression vectors, host cells and methods for producing BasiB13 polypeptides using recombinant methods. Also claimed is a vaccine composition comprising a BasiB13 polypeptide, an immunogenic

CC fragment of a BsrB13 polypeptide, or a polypeptide having at least
CC 85% amino acid sequence identity to BsrB13, or comprising a
CC polynucleotide encoding such a polypeptide. A claimed method of
CC diagnosing a Moraxella infection involves identifying a BsrB13
CC polypeptide or antibody. A claimed therapeutic composition useful
CC in treating humans with M. catarrhalis infection comprises at least
CC 1 antibody directed against a BsrB13 polypeptide. BsrB13
CC polypeptides also have utility in raising specific antibodies
CC and in screening for antibacterial drugs.
XX
Sequence 224 AA;
SQ

Query Match	13.3%	Score 111;	DB 22;	Length 224;
Best Local Similarity	33.3%	Pred. No. 0.00097;		
Matches	37;	Conservative	15;	Mismatches 25;

	Accession	Size	Inserts	Gaps
Q7	7 GSSLIIIVFLVGCACNFRSEVGAATG-----AVVGVGAOLFGRKSGRVSATIGG	58	11	11
Db	7 GVVLIIASMLALCANTGT-----TGNGTGFEGANVNAVIGAAVAGAL-----GGTAISATG	60	11	11
Q7	59 -----AVLGGIGSGITGSMDDDKTKLNLQSEKVKAGAVTRMNPDTG	103	11	11
Db	61 EKTGRDAILCAAVGAAGAATMERQAR-----QLEQMGQGTGVTHDDITGN	107	11	11

```

RESULT 7
AAV34487
CID      AAV34487  standard; Protein; 223 AA.
xx

```

AAC	AAY34487;	
XX		
DT	25-AUG-1999	(first entry)
XX		

Porphyromonas gingivalis protein PG3.

W vaccine; antigenic.

Porphyromonas gingivalis.

17-JUN-1999

10-DEC-1998; 98WO-AU01023.

04-AUG-1998; 98AU-0005028.
10-DEC-1997

31-DEC-1997; 97AU-0001182.

10-MAR-1998; 98AU-0002264

23-APR-1998; 98AU-0002911.
98AIT-0003128

05-MAY-1998; 98AU-0003338.
22-MAY-1998; 98AT-0003554

29-JUL-1998; 98AU-0004917.

(CSLC-) CSL LTD.

WPI: 1999-385613/32.
N-PSDB: AAX91705.

Antigenic Porphyromonas gingivalis peptides for preventing gingivitis
Claim 1; Page 469; 588pp; English.

AAV9136 to AAV91801 encode two hundred and sixty six antigenic Porphyromonas gingivalis (Pg) polypeptide sequences given in AAV4318 to AAV34583. AAV91802 to AAV91989 represent PCR primers used in the

CC Isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.

XX Sequence 223 AA;

Query Match 12.3%; Score 102.5; DB 20; Length 223;

Best Local Similarity 34.7%; Pred. No. 0.0079;

Matches 25; Conservative 13; Mismatches 31; Indels 3; Gaps 2;

OY 8 SSIITVFVLCGCAQNFSSQEVGAATGAVGVAAGOLFSGSGRVSMA-ATGAVLGLIG 66

DB 8 ASYLAVALVFAGCGGLN--NMAKGGLIGAGVGGAIGAGVGNAGNTAVGATVGAAGGAG 65

OY 67 SKIGOSMDQDK 78

DB 66 ALIGKMKDKOKK 77

RESULT 8

AAV34362

ID AAV34362 standard; Protein; 230 AA.

XX AAV34362;

DT 25-AUG-1999 (first entry)

DE Porphyromonas gingivalis protein PG3.

KM Porphyromonas gingivalis; PG; periodontal disease; gingivitis;

KW vaccine; antigenic.

OS Porphyromonas gingivalis.

PN WO9929870-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98WO-AU01023.

PR 04-AUG-1998; 98AU-0005028.

PR 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

XX (CSLC-) CSL LTD.

PI Agius CF, Barr IG, Hocking DM, Margetts MB, Patterson MA;

PI Ross BC, Rothel LJ, Webb EA;

XX WPI; 1999-385613/32.

DR N-PSDB; AAX91580.

XX Antigenic Porphyromonas gingivalis peptides for preventing

PT gingivitis

XX Claim 1; Page 325-326; 588pp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic

CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to

CC AAY34583. AAX91802 to AAX91869 represent PCR primers used in the

CC isolation of the PG polypeptides. The PG polypeptides have antibacterial

CC activity with a vaccine mechanism of action. The PG polypeptides can be

CC used as vaccines especially against Porphyromonas gingivalis. Probes can

CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.

XX Sequence 230 AA;

Query Match 12.3%; Score 102.5; DB 20; Length 230;

Best Local Similarity 34.7%; Pred. No. 0.0082;

Matches 25; Conservative 13; Mismatches 31; Indels 3; Gaps 2;

OY 8 SSIITVFVLCGCAQNFSSQEVGAATGAVGVAAGOLFSGSGRVSMA-ATGAVLGLIG 66

DB 15 ASYLAVALVFAGCGGLN--NMAKGGLIGAGVGGAIGAGVGNAGNTAVGATVGAAGGAG 72

OY 67 SKIGOSMDQDK 78

DB 73 ALIGKMKDKOKK 84

RESULT 9

AAR05799

ID AAR05799 standard; Protein; 154 AA.

XX AAR05799;

DT 31-AUG-1990 (first entry)

DE PBOMP-2 gene prod.

KM Outer membrane proteins; PBOMP-2; immunisation;

KW Praxis Biologics Outer Membrane Protein.

OS Haemophilus influenzae.

PN WO9002357-A.

PD 22-MAR-1990.

PF 31-AUG-1989; 89WO-US03779.

PR 01-SEP-1988; 88US-0239572.

PR 21-AUG-1989; 89US-0396572.

XX (PRAXIS) PRAXIS BIOLOGICS IN.

PI Anilionis A, Seid RC, Deich RA, Zlotnick GW, Green BA;

DR WPI; 1990-115815/15.

DR N-PSDB; AAO03870.

XX Outer membrane protein epitopes of Haemophilus influenzae- used in

PT the prodn. of antibodies. In vaccines and for prodn. of reagents for

PT diagnosis.

XX Disclosure; Fig 15; 164pp; English.

PS The PBOMP proteins were isolated from a PBOMP-enriched insoluble

CC cell wall fraction from physically disrupted cells of H. influenzae

CC and then solubilising the PBOMP from the cell wall fraction by heating

CC in the presence of a detergent or digesting the cell wall fraction

CC with lysozyme, opt. in the presence of a detergent. The genes

CC encoding the PBOMP proteins were isolated by screening a DNA library

CC with an oligonucleotide probe based on the amino acid sequence of the

CC PBOMP protein, or using antibodies to PBOMP.

XX Sequence 154 AA;

Query Match 12.1%; Score 101; DB 11; Length 154;

Best Local Similarity 39.6%; Pred. No. 0.0071;

Matches 21; Conservative 9; Mismatches 23; Indels 0; Gaps 0;

OY 29 VGAATGAVGVAGOLFSGSGRVSMAIGAVLGLIGSKIGOSMDQDKIKL 81

CC results in better recovery of protein at higher purity (70-99%),
CC is easy to scale up, and the fibres are spun in an environmentally
CC benign solution reducing hazardous waste accumulation and cost.

XX Sequence 528 AA;

Query Match 10.6%; Score 88; DB 22; Length 528;

Best Local Similarity 48.9%; Pred. No. 0.87; Mismatches 19; Indels 2; Gaps 1;

Matches 23; Conservative 3; Mismatches 19; Indels 2; Gaps 1;

QY 27 QEVGAATGAVGVAGQ-LFGKSGRVSMAIGAVLGIGSKIGQ 71

Db 320 QGAGAAAAAGAGAGGCGGAGAGAAAAAGAGAGCGGCGAGQ 366

RESULT 12

ABB69847 standard; Protein; 147 AA.

XX ABB69847;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 36333.

KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.

OS Drosophila melanogaster.

PN MO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001MO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABLI3950.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX Disclosure; SEQ ID NO 36333; 21pp + Sequence Listing: English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 147 AA;

Query Match 10.5%; Score 87.5; DB 22; Length 147;

Best Local Similarity 27.5%; Pred. No. 0.19; Mismatches 39; Gaps 7;

Matches 36; Conservative 18; Mismatches 38; Indels 39; Gaps 7;

QY 36 VVGAVAGLFGKSGRVSMAIGAVLGIGSKIGQMDQ---DKTKLN 82

Db 11 VTSSISS--IGLGNRRFSLSNQSSSQSGIGGAIPEGIVSGAQQQQQVPRPQPLQ 68

QY 83 OSLEKVKAGVTRMRNPDFTGNSYSVEPVTRYORNKORROQYCR-EFOOKAMIAQOK 141

Db 69 QSLPQQQQQ-----00000001000HGHAPLQ000AST-GSGLG 106

QY 142 IYGTACPPDG 152

Db 107 YGTAPR-PTG 116

RESULT 13

ABB66232 standard; Protein; 2309 AA.

XX ABB66232;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 25488.

KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.

OS Drosophila melanogaster.

PN MO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001MO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABLI0335.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX Disclosure; SEQ ID NO 25488; 21pp + Sequence Listing: English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 2309 AA;

Query Match 10.4%; Score 86.5; DB 22; Length 2309;

Best Local Similarity 27.4%; Pred. No. 8.5; Mismatches 55; Indels 17; Gaps 5;

Matches 34; Conservative 18; Mismatches 55; Indels 17; Gaps 5;

QY 22 QNFSROEVGATGA---VVGAVAGLFGKSGRVSMAIGAVLGIGSKIGQMDQDK 78

Db 2124 QQQQOQVGVGGNGSPSMALGRCGAVGSGSGN-----GGGGGAGAGSGVGGGNGV 2178

QY 79 IKLNQSLKVKAGVTRMRNPDFTGNSYSVEPVTRYORNKORROQYCRFPQOKAMIA 137

Db 2179 GSVGQS-----GGGGGGRVPRPIORPNVYPOHP-----000000000000REDAAAAAVAA 2230

OY 138 OKOE 141
ID 1:
DB 2231 ORAQ 2234

RESULT 14

ABB61254
ID ABB61254 standard; Protein; 1251 AA.

XX ABB61254;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 10554.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL05357.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Disclosure; SEQ ID NO 10554; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX sequences (ABL01840-ABL16175) and the encoded proteins

XX (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1251 AA:

Query Match 10.1%; Score 84; DB 22; Length 1251;
Best Local Similarity 22.7%; Pred. No. 7.1;
Matches 35; Conservative 28; Mismatches 67; Indels 24; Gaps 5;

OY 1 MRGCLGSSLIITSVFLVCAQNFNRQEVGAATGAVGVAGGLFGSGSRVSMATGAV 60

DB 915 MGGMGOMGOMGOMGOMGOMGOMGOMGOMGOMGOMGOMGOMGOMGOMGOMGOM 971

OY 61 LGLLIGSK--IGOSMDQODIKLNOSLEKVRAGVTRMRNPDTGNSVSEPVRTYQRYNK 118

DB 972 PGQMMGRGRGLNQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 1022

OY 119 QERRQOYCFREFQOKAMTAGCKOETIGTACPOPDG 152

DB 1023 OKKPRYFVAMFDYDP-----STMSPNPDG 1046

RESULT 15

AA11028
ID AA11028 standard; Protein; 116 AA.

XX AA11028;

XX 08-JUN-1999 (first entry)

XX H. pylori ORF 01cp20708_10628177_c2.50 secreted protein.

XX Vaccine; probe; diagnostic; ORF; cell envelope protein;

XX secreted protein; cytoplasmic protein; cellular protein.

XX Helicobacter pylori.

XX WO9824475-A1.

XX 11-JUN-1998.

XX 05-DEC-1997; 97WO-US22104.

XX 14-JUL-1997; 97US-0891928.

XX 05-DEC-1996; 96US-0759625.

XX 25-MAR-1997; 97US-0823745.

XX (ASTR) ASTRA AB.

XX Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;

XX WPI; 1998-333051/29.

XX N-PSDB; AAX30557.

XX New isolated Helicobacter pylori nucleic acids - used to develop

XX products for the diagnosis, prevention and treatment of infection by

XX H. pylori and other Helicobacter species

XX Claims 37, 41; Page 190-191; 339pp; English.

XX Recombinant or substantially pure preparations of H. pylori polypeptides

XX are disclosed, together with the nucleic acids encoding them. In all,

XX 97 ORFs are shown. The proteins are variously cell envelope proteins,

XX cytoplasmic proteins, secreted proteins or other cellular proteins.

XX Vaccines containing the nucleic acids or proteins are claimed, as are

XX probes containing at least 8 nucleotides from the nucleic acid

XX sequences. The vaccines are useful for treating or reducing the risk of

XX H. pylori infections, and the probes can be used diagnostically for

XX detecting the presence of Helicobacter in a sample. The products are

XX also of use in screening for compounds having the ability to interfere

XX with the H. pylori life cycle or to inhibit H. pylori infection.

XX Sequence 116 AA:

Query Match 10.0%; Score 83; DB 19; Length 116;
Best Local Similarity 26.7%; Pred. No. 0.42;
Matches 31; Conservative 16; Mismatches 51; Indels 18; Gaps 4;

OY 1 MRGCLGSSLIITSVFLVCAQNFNRQEVGAATGAVGVAGGLFGSGSRVSMATGAV 60

DB 11 MSLKRTFSCVYVLCGMVNAVAVAGPKTEARGELGKTVGAVGFMVDKMGCR--VGCAI 67

OY 61 LGLLIGSKITGOSMDQODIKLNOSLEKVRAGVTRMRNPDTGNSVSEPVRTYQRYNK 116

DB 68 -GGYISSEVGRD--EDYIRGVDRREPQNNKEPQPR-----ETIRDFYDY 108

Search completed: October 27, 2002, 11:09:11
Job time : 21.411 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 15:09:08 : Search time 1529.81 Seconds
(without alignments)
6775.784 Million cell updates/sec

Title: US-09-677-374-5
Perfect score: 768
Sequence: 1 atgtcagtgatctacaa.....aggtgattagcaccgaaaaa 768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlinu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_hic:*
9: gb_estli:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_huv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43.4	5.7	551	9	AU037653 AU037653
2	43	5.6	1100	12	CNS00FCD AU070512 Drosophila
3	42.4	5.5	470	12	AZ799648 2M057B14
4	42.2	5.5	976	12	CNS04E5M AL286627 Tetradon
5	42	5.5	238	10	BG240789 OVI_38 CO
6	42	5.5	273	12	AQ260501 CITR1-El
7	42	5.5	402	10	BG356895 OV2_11-DO
8	42	5.5	419	10	BE353894 DG1_11-El
9	42	5.5	428	10	BF176742 EM1_4-F01
10	42	5.5	446	10	BG052290 RH12_12-
11	42	5.5	500	10	BF586874 FMI_31-EO
12	42	5.5	507	10	BG053591 RH12_11-
13	42	5.5	531	10	BG053014 RH12_16-
14	42	5.5	537	10	BG713844 EM1_20-CO
15	42	5.5	561	10	BF587805 FMI_40-G0
16	42	5.5	577	10	BG673858 EM1_20-CO
17	42	5.5	582	10	BE356763 DG1_12-B1

18	42	5.5	591	10	BE356800 DG1_12-El
19	42	5.5	597	10	BG102589
20	42	5.5	598	9	AM672446 LG1_359_L0
21	42	5.5	600	10	BE360868
22	42	5.5	620	10	BF481524 FMI_21-B0
23	42	5.5	649	10	BE355895
24	41.8	5.4	238	9	AA352399 EST60642
25	41.8	5.4	474	10	BM031893
26	41.6	5.4	321	9	AM021656
27	41.6	5.4	361	9	AA653164
28	41.6	5.4	459	10	BI493124
29	41.6	5.4	509	12	AZ447316
30	41.6	5.4	556	10	BE395015
31	41.6	5.4	597	10	BI492596
32	41.6	5.4	671	9	AM328753
33	41.6	5.4	742	10	BF792098
34	41.6	5.4	878	10	BM449796
35	41.6	5.4	958	12	CNS001FF
36	41.6	5.4	1058	10	BM478067
37	41.6	5.4	1079	10	BF792265
38	41.6	5.4	1201	12	CNS0167M
39	41.2	5.4	702	10	W27594
40	41.2	5.4	737	10	BM166352
41	41	5.3	928	12	CNS00DKY
42	40.6	5.3	942	12	CNS018GS
43	40.4	5.3	548	10	BM328366
44	40.4	5.3	609	12	AZ421257
45	40.4	5.3	856	12	A0740253 HS_5505_A

ALIGNMENTS

RESULT 1
LOCUS AU037653 551 bp mRNA linear EST 29-MAR-1999
DEFINITION AU037653 Dictyostellium discoideum SS (H.Urushihara) Dictyostellium
discoideum cDNA clone SS138, mRNA sequence.

ACCESSION AU037653
VERSION AU037653.1 GI:3984406
KEYWORDS EST.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum.
REFERENCE 1 (bases 1 to 551)
Mori, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
Yoshino, R., Mitra, B.N., Pl, M., Sato, T., Takemoto, K., Yasukawa, H.,
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
The Dictyostellium development cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development
DNA Res. 5 (6), 335-340 (1998)

JOURNAL MEDLINE
CONTACT: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402huesakura.cc.tsukuba.ac.jp
PROJECT = 'Dictyostellium discoideum cDNA project in Japan'.
Location/Qualifiers
1. 551
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="SSE138"
/clone="Dictyostellium discoideum SS (H.Urushihara)"
/dev_stage="slug"

BASE COUNT 266 a 63 c 39 g 183 t
ORIGIN

Query Match 5.7%; Score 43.4; DB 9; Length 551;
Best Local Similarity 52.5%; Pred. No. 3;
Matches 95; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```

QY      32 CAGCACAACAATCATATTACACATATCAAAATTTACTACACATCTGCAGTGATT 91
      34 CAACTATACAAATTAATTTAAATATCAATAAAAATCGATTAATCAAAAAATTTAAAT 93
QY      92 TAAATTTAATGACGTAAGATTAGATTATTATTCACACAGTGTGTACACAGCACA 151
      94 CAAATTCACATATTTTAAAAAAGAAATTTTAAATCAATGTTACTTCATCAATCAAA 153
QY      152 CTTTCTGTGTGACATGCTGTCATTTATAGGAATATGCTATGTTGATTAACACTAGCA 211
      154 ATTCAATTAATTAATTGCAACTTTAATTTTACCAAAAAATGATGATTAATATAATA 213
QY      212 A 212
      214 A 214

RESULT 2
CNS00FCD 1100 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BAC3D18 of RPCI-98 library from Drosophila melanogaster (fruit
AL070512 fly), genomic survey sequence.
ACCESSION AL070512.1 GI:4950453
VERSION GSS.
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1100)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw.sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Source
1. 1100
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BAC3D18"
/note="end : TET3"
BASE COUNT 356 a 100 c 111 g 137 t 396 others
ORIGIN
Query Match 5.6%; Score 43; DB 12; Length 1100;
Best local Similarity 33.7%; Pred. No. 5;
Matches 83; Conservative .6; Mismatches 157; Indels 0; Gaps 0;
QY 25 AACCAATCAGCACAACAATCAATTTACACCAATATCAACATCTGC 84
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 104 AAAAAAAAAAAAAAAAAAAAAAAAAANACATNNNNATATTNNACAATNNANN 163
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 85 ACGGATTTAATTTAATGACGTAAGATTAGATATTATTACACAAGTGATGTACACAA 144
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

```

```

Db 164 CGASGGAANAANAAGGAANNANNAANAANAHAAYAAANATTAAMAAAAA 223
QY 145 GGACCAACTTTCTGTGTGACCATCTGTCATTTATAGAAATAGCTATGTTGATPAC 204
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 224 CAAAGCATTTTANNNNNNTNNNTNNNTNNNNANANNNANNNANNNNNANNN 283
QY 205 ACTAGCAAGTGACAGCAAACTTGTTAAAGAAACAGCAGCCCAACATCACTATGAT 264
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 284 NAAAAAANNNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 343
QY 265 ACATAT 270
      11 11 11
Db 344 AAKAT 349

RESULT 3
A2799648 470 bp DNA linear GSS 16-FEB-2001
LOCUS 2M0057B14F Mouse 10kb plasmid UGCLM library Mus musculus genomic
DEFINITION clone UGCG2M0057B14 F, DNA sequence.
ACCESSION A2799648
VERSION A2799648.1 GI:12950975
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 470)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beecorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
Plasmid Inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 1000 Std Error: 0.00
Plate: 0057 row: B column: 14
Seq primer: CGTTGTAACAGCAGCGCCAGT
Class: plasmid ends
High quality sequence stop: 470.
FEATURES
Source
1. 470
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCG2M0057B14"
/clone_lib="Mouse 10kb plasmid UGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: PWD42ny. Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g14732114|9b|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to

```


Db	246	AA	247
RESULT 7			
LOCUS	RG356895	402 bp	mRNA linear EST 06-MAR-2001
DEFINITION	OV2.11.D01.g1_A002 Ovary 2 (OV2)	Sorghum bicolor	CDNA, mRNA
ACCESSION	RG356895		
VERSION	RG356895.1	GI:13238881	
KEYWORDS	EST.		
SOURCE	Sorghum.		
ORGANISM	Sorghum bicolor		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.		
AUTHORS	1 (bases 1 to 402) Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt ,L.H.		
TITLE	An EST database from Sorghum: ovaries of varying immature stages		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Cordonnier-Pratt MM The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmprratt@uga.edu Seq primer: PolyTmix High quality sequence start: 47 High quality sequence stop: 391 POLYA-No.		
FEATURES			
Source	Location/Qualifiers		
	1..402		
	/organism="Sorghum bicolor"		
	/db_xref="taxon:4558"		
	/clone_lib="Ovary 2 (OV2)"		
	/note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from Lambda zap II; Site.1: XhoI; Site.2: EcoRI; The library was made from poly A RNA in the cloning vector Lambda zap II. Clones to be sequenced were prepared by mass excision."		
BASE COUNT	107 a 82 c 117 g 96 t		
ORIGIN			
Query Match	5.5%; Score 42; DB 10; Length 402;		
Best Local Similarity	59.0%; Pred. No. 5.9;		
Matches	72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;		
QY	345 GCGCCAGACTTTCAGACGCCAGGAAGTTGGCGCGGCCACCGCGTGTGGCGCGTGTG 404		
DB	140 CGGCGAGAGCTGCACCGCAGCAGGAGTGCAGCAGATGATCCTCGAGGCTGACCTCATGG 199		
QY	405 TGCCGGCCAGCTGTTCGTTAAGGCTGTGCTCGTGTGCATGCGCATGCGCCGTCGCGT 464		
DB	200 CGAGCGCCAGTCACTATGACAGAGTGTGTTAAGTTATGATGCGCAAGTGAAGACCGGT 259		
QY	465 TC 466		
DB	260 CC 261		
RESULT 8			
LOCUS	BR355894	419 bp	mRNA linear EST 20-JUL-2000
DEFINITION	DB1.11.E11.g1_A002 Dark Grown 1 (DG1)	Sorghum bicolor	CDNA, mRNA
ACCESSION	BR355894		
VERSION	BR355894.1	GI:9297451	
KEYWORDS	EST.		
SOURCE	Sorghum.		
ORGANISM	Sorghum bicolor		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 419)
Cordonnier-Pratt, M.-M., Gingle, A., Matsuda, C., Sudman, M. and Pratt
REFERENCE
AUTHORS

TITLE	JOURNAL	COMMENT
An EST database from Sorghum: dark-grown seedlings	Unpublished (2000)	Contact: Cordonnier-Pratt MM

Email: mmparratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: POLYTMix
High quality sequence start: 16
High quality sequence stop: 336
POLYA=No.

FEATURES	Location/Qualifiers
source	1. .419

```

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DG1)"
/notes="Organ: 5-day-old dark-grown seedlings; Vector:
lambda Zap; Site_1: XhoI; Site_2: EcoRI. The library was
made from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT      102 a      82 c      116 g      119 t
ORIGIN

```

```
Query Match      5.5%; Score 42; DB
```

Best Local Similarity 59.0%; Pred. NO. 6;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0

345 CCCCCAGAACTTCAGGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGCGGTTG 404

Db 78 CCGGCGAAGACGACCTAGGAGGTGCACAGATGATCCGTGAGGCTGACGTGCATGG 137

Qy 405 TCCCGGCCAGCTGTTCCGGTAAAGGCTCTGCTGTGTCGATGGCCATCGCGGTGGGT 464

Db 138 GCACGGCCAGATCAACTATGAGAGAGTTGTATAGGTTATGATGGCCAACTGAGGAGCGGT 197

Qy 465 TCC 466

Db 198 CC 199

RESULT	9
LOCUS	Bf176742
DEFINITION	Bf176742 428 bp mRNA linear EST 31-OCT-2001 EM_4_F01.g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION	Bf176742
VERSION	Bf176742.1 GI:11064652
KEYWORDS	EST.
SOURCE	sorghum.
ORGANISM	Sorghum bicolor

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Rosaceae: PACC clade: Panicoideae: Andropogoneae: Sorghum.
1 (Passes 1 to 428)
Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
An EST database from Sorghum: developing embryos
Unpublished (2000)
Contact: Cordonnier-Pratt MM

Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805

Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

High quality sequence start: 32
High quality sequence stop: 427
POLYA=NO.

FEATURES	Location/Qualifiers
source	1. .428

```

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/notes="Organ: Embryos germinated for 24 hr. Vector:
pBluescript II from lambda Zap II. Site_1: XhoI; Site_2:
EcoRI. The library was made from poly-A RNA in the cloning
vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT      103 a      80 c      119 g      126 t

```

ORIGIN

Query Match	5.5%;	Score 42;	DB 10;	Length 428
Best Local Similarity	59.0%;	Pred. No. 6.1;		

Matches	72;	Conservative	0;	Mismatches	50;	Indels	0;	Gaps	0
---------	-----	--------------	----	------------	-----	--------	----	------	---

[illegible]

Db	178	CC	179
RESULT	10		
LOCUS	BG052290		
DEFINITION	BG052290	446 bp	mRNA
	RH1Z2_12.G04.g1_A003	linear	EST 25-JAN-2001
	sequence.		
ACCESSION	BG052290		
VERSION	BG052290.1		
KEYWORDS	EST.		
SOURCE	Sorghum		propiquinum.
ORGANISM	Sorghum		propiquinum

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eurariopsis, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Sorghum.
1 (bases 1 to 446)
Cordonnier-Pratt, M. M., Gingle, A., Paterson, A., Sudman, M. and Pratt
L. H.
An EST database from Sorghum: Sorghum proproinquum rhizomes
Unpublished (2000)
Contact: Cordonnier-Pratt MM

Department Of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Seq primer: PolyTmix
High quality sequence start: 11
High quality sequence stop: 440
POLYA-NO.

```
FEATURES
  source      Location/Qualifiers
1.  .446      1.  .446
                /organism="Sorghum proflinquinum"
                /db_xref="taxon:132711"
                /clone_lib="Rhizome2 (RHIZ2)"
                /note="Organ: Rhizomes; Vector: pBluescript II from Lambda"
```

```
Location/Qualifiers
1..446
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RHIZ2)"
/note="Organ: Rhizomes; Vector: pBluescript II from Lambda"
```

```
Location/Qualifiers
1..446
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RHIZ2)"
/notes="Organ: Rhizomes; Vector: pBluescript II from Lambda"
```

```
Location/Qualifiers
1..446
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RHIZ2)"
/notes="Organ: Rhizomes; Vector: pBluescript II from Lambda"
```

Zap II: Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision.

BASE COUNT
ORIGIN

101 a 81 c 124 g 140 t

Query Match

Best Local Similarity 5.5%; Score 42; DB 10; Length 446;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 345 CCCCAGAACTTCAGCCGCCAGAGTTGGCGGCCACCGGTGGTGGGGGTGT 404
DB 51 CGCGGAGAAAGCTGACCGACGAGAGGTGCGACGATGATCCGTCGAGCTGATGG 110
QY 405 TCGCGCCAGCTGTTCGTTAAAGCTCTGCTGTGTCGATGCGCATGCGCGTGG 464
DB 111 CGACGGCCAGATCAACTATGACAGTTTGTATGATGATGAGGAGTGAAGCGGT 170
QY 465 TC 466
DB 171 CC 172

RESULT 11

BF586874

LOCUS

DEFINITION

FM1_31.E06.g1.A003 Floral-Induced Meristem 1 (FM1) Sorghum

ACCESSION

BF586874

VERSION

BF586874.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

TITLE

Unpublished (2000)

Contact: Cordomier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmp@prattuga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence

is 20.

Seq primer: PolyTMIx

High quality sequence start: 8

High quality sequence stop: 495

POLYA-No.

Location/Qualifiers

1..500

/organism="Sorghum propinquum"

/db_xref="taxon:132711"

/clone_lib="Floral-Induced Meristem 1 (FM1)"

/note="Organ: Floral-Induced Meristems; Vector: pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested. The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT

ORIGIN

111 a 94 c 141 g 154 t

Query Match
Best Local Similarity 5.5%; Score 42; DB 10; Length 500;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 345 CCCCAGAACTTCAGCCGCCAGAGTTGGCGGCCACCGGTGGTGGGGGTGT 404
DB 89 CGCGGAGAAAGCTGACCGACGAGAGGTGCGACGATGATCCGTCGAGCTGATGG 148
QY 405 TCGCGCCAGCTGTTCGTTAAAGCTCTGCTGTGTCGATGCGCATGCGCGTGG 464
DB 149 CGACGGCCAGATCAACTATGACAGTTTGTATGATGATGAGGAGTGAAGCGGT 208
QY 465 TC 466
DB 209 CC 210

RESULT 12

BG053591

LOCUS

DEFINITION

RH122_11.C03.g1.A003 Rhizome2 (RH122) Sorghum propinquum cDNA, mRNA

ACCESSION

BG053591

VERSION

BG053591.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

TITLE

Unpublished (2000)

Contact: Cordomier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmp@prattuga.edu

Seq primer: T7

High quality sequence start: 95

High quality sequence stop: 507

POLYA-yes.

Location/Qualifiers

1..507

/organism="Sorghum propinquum"

/db_xref="taxon:132711"

/clone_lib="Rhizome2 (RH122)"

/note="Organ: Rhizomes; Vector: pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT

ORIGIN

109 a 92 c 142 g 164 t

Query Match

Best Local Similarity 5.5%; Score 42; DB 10; Length 507;
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 345 CCCCAGAACTTCAGCCGCCAGAGTTGGCGGCCACCGGTGGTGGGGGTGT 404
DB 54 CGCGGAGAAAGCTGACCGACGAGAGGTGCGACGATGATCCGTCGAGCTGATGG 113
QY 405 TCGCGCCAGCTGTTCGTTAAAGCTCTGCTGTGTCGATGCGCATGCGCGTGG 464
DB 114 CGACGGCCAGATCAACTATGACAGTTTGTATGATGATGAGGAGTGAAGCGGT 173
QY 465 TC 466
DB 174 CC 175

REFERENCE	1 (bases 1 to 537)					
AUTHORS	Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.					
TITLE	An EST database from sorghum: developing embryos					
JOURNAL	Unpublished (2000)					
COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmprratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20. Seq primer: PolyTMix High quality sequence start: 62 High quality sequence stop: 536 POLYA-No.					
FEATURES						
SOURCE	Location/Qualifiers 1..537 /organism="Sorghum bicolor" /db_xref="taxon:4558" /clone_id="Embryo 1 (EM1)" /note="Organ: Embryos germinated for 24 hr; Vector: pbuescript II from lambda zap II; Site.1: XhoI; Site.2: EcoRI; The library was made from poly-A RNA in the cloning vector Lambda ZAP II. Clones to be sequenced were prepared by mass excision." BASE COUNT 118 a 98 c 152 g 169 t ORIGIN					
Query Match	5.5%; Score 42; DB 10; Length 537;					
Best Local Similarity	59.0%; Pred. No. 6.6;					
Matches	72; Conservative 0; Mismatches 50; Indels 0; Gaps 0:					
DQ	345	CGCCAGACACTTCAGGCCGCAAGAATGGGCGGCACCOCGTGGCTGGTGGCGGT	404			
DQ	92	C GGCGAGAGCTGACCGCAGAGAGTGACGATGATCCGTGAGCTGACGTGATGG	151			
DQ	405	TGCGGCGCACGCTGTTCGTTAAAGGCTGCGTGTGATGATGCCATCGCGTGCGGT	464			
DQ	152	CGAGCGCCAGATCAACTGTGAAGATTGTTAAGTTATGATGCGCAAGTAGAGCGGT	211			
DQ	465	TC 466				
	-					
DQ	212	CC 213				
RESULT 15						
BFS87805	BFS87805 561 bp mRNA linear EST 12-DEC-2000					
LOCUS	FMI_40.G05.q1.A003 floral-induced Meristem 1 (FMI) Sorghum					
DEFINITION	propinquum cdNA, mRNA sequence.					
ACCESSION	BF587805					
VERSION	BF587805.1 GI:11680115					
KEYWORDS	EST.					
SOURCE	Sorghum propinquum.					
ORGANISM	Sorghum propinquum Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC clade; Panicoideae; Andropogoneae; Sorghum. 1 (bases 1 to 561) Cordonier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt, L.H. An EST database from Sorghum: floral-induced meristems Unpublished (2000) Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email:mmprratt@uga.edu					
TITLE						
JOURNAL						
COMMENT						

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: PolyTMix
High quality sequence

High quality sequence start: 81
High quality sequence stop: 561

POLYA=NO.

POLYA=NO.

FEATURES

SOURCES	Location/Qualifiers
source	1. .561

1. .561

/organism="Sorghum propinquum"
/db xref="taxon:132717"

/db_xref="taxon:132711"

```
/clone_lib="Floral-Induced Meristem 1 (FMI)"
```

/note="Organ: Floral-induced meristems; Vector:

Pbluescript II from Lambda Zap II; Site_1: XhoI; Site_2:

ECORI; mature plants were placed in a growth chamber for

15 days with 16 hr darkness and 8 hr light (flowering 1

induced by short-day conditions); 16 days after being

returned to the greenhouse under natural long days during

late April/early May. Meristems were harvested The

rate April/early May, meristems were harvested. The library was made from poly-A RNA in the cloning vector

Idially was made from poly-A RNA in the cloning vector

mass excision."

BASE COUNT	140 a	111 c	160 g	150 t
ORIGIN				

ORIGIN

Query Match

5.58; Score 42; DB 10; Length 561;

Best Local Similarity 59.08; Pred. No. 6.8;

Matches	72;	Conservative	0;	Mismatches	50;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

345 CGCCGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGGGTTGTGGCGGTGT 404

Db 202 CGCGGAGAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGC 261

405 TCCCGGCCAGCTGTTCCGCTAAAGGCTCTGGTCTGTGTCATGACCTATCCGGCGTCGCT 464

[illegible]

QY	465	TC	466
----	-----	----	-----

Db 322 CC 323

Search completed: October 27, 2002, 18:34:59
Job time : 1533.81 secs

Job time : 1533.81 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 16:06:08 : Search time 56.152 Seconds
(without alignments)
3359.570 Million cell updates/sec

Title: US-09-677-374-5

Perfect score: 768

Sequence: 1 atcgacgttgaaattctacaa.....aggtgattagcaccgaaaaa 768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/2/1na/5A_COMB.seq:*
2: /cgn2_6/prodata/2/1na/5B_COMB.seq:*
3: /cgn2_6/prodata/2/1na/6A_COMB.seq:*
4: /cgn2_6/prodata/2/1na/6B_COMB.seq:*
5: /cgn2_6/prodata/2/1na/PCRTUS_COMB.seq:*
6: /cgn2_6/prodata/2/1na/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	270.2	35.2	486	1 US-08-048-164A-1	Sequence 1, Appl
2	270.2	35.2	486	1 US-08-048-164A-3	Sequence 3, Appl
3	270.2	35.2	486	1 US-08-460-462-1	Sequence 1, Appl
4	270.2	35.2	486	1 US-08-460-462-3	Sequence 3, Appl
5	270.2	35.2	486	1 US-08-460-457-1	Sequence 1, Appl
6	270.2	35.2	486	1 US-08-460-457-3	Sequence 3, Appl
7	270.2	35.2	486	1 US-08-460-458-1	Sequence 1, Appl
8	270.2	35.2	486	1 US-08-460-458-3	Sequence 3, Appl
9	270.2	35.2	486	2 US-08-460-455-1	Sequence 1, Appl
10	270.2	35.2	486	2 US-08-460-455-3	Sequence 3, Appl
11	270.2	35.2	486	2 US-08-330-394A-1	Sequence 1, Appl
12	270.2	35.2	486	2 US-08-330-394A-3	Sequence 3, Appl
13	270.2	35.2	486	3 US-09-006-636-5	Sequence 5, Appl
14	270.2	35.2	486	4 US-09-006-632-5	Sequence 21, Appl
15	270.2	35.2	486	4 US-09-277-716-21	Sequence 9, Appl
16	70.2	9.1	1482	4 US-09-198-956-9	Sequence 11, Appl
17	70.2	9.1	1482	4 US-09-198-955A-11	Sequence 31, Appl
18	40	5.2	2645	4 US-08-560-780-31	Sequence 31, Appl
19	40	5.2	2645	4 US-09-073-898-31	Sequence 31, Appl
20	40	5.2	2645	4 US-09-073-898-31	Sequence 31, Appl
21	39.4	5.1	2004	1 US-08-471-033-6	Sequence 6, Appl
22	39.4	5.1	2004	2 US-08-471-044-6	Sequence 6, Appl
23	39.4	5.1	2004	2 US-08-463-483A-6	Sequence 6, Appl
24	39.4	5.1	2004	2 US-08-471-046A-6	Sequence 6, Appl
25	39.4	5.1	2004	2 US-08-470-566B-6	Sequence 6, Appl
26	39.4	5.1	2004	2 US-08-469-334-6	Sequence 6, Appl
27	39.4	5.1	2004	3 US-09-300-529-6	Sequence 6, Appl

28	39.4	5.1	2655	1 US-08-471-033-4	Sequence 4, Appl
29	39.4	5.1	2655	2 US-08-471-044-4	Sequence 4, Appl
30	39.4	5.1	2655	2 US-08-463-483A-4	Sequence 4, Appl
31	39.4	5.1	2655	2 US-08-471-046A-4	Sequence 4, Appl
32	39.4	5.1	2655	2 US-08-470-566B-4	Sequence 4, Appl
33	39.4	5.1	2655	2 US-08-469-334-4	Sequence 4, Appl
34	39.4	5.1	2655	3 US-09-300-529-4	Sequence 4, Appl
35	39.4	5.1	4041	1 US-08-471-033-22	Sequence 22, Appl
36	39.4	5.1	4041	2 US-08-471-044-22	Sequence 22, Appl
37	39.4	5.1	4041	2 US-08-463-483A-22	Sequence 22, Appl
38	39.4	5.1	4041	2 US-08-471-046A-22	Sequence 22, Appl
39	39.4	5.1	4041	2 US-08-470-566B-22	Sequence 22, Appl
40	39.4	5.1	4041	2 US-08-469-334-22	Sequence 22, Appl
41	39.4	5.1	4041	3 US-09-300-529-22	Sequence 22, Appl
42	39.4	5.1	6049	1 US-08-471-033-1	Sequence 1, Appl
43	39.4	5.1	6049	2 US-08-471-044-1	Sequence 1, Appl
44	39.4	5.1	6049	2 US-08-463-483A-1	Sequence 1, Appl
45	39.4	5.1	6049	2 US-08-471-046A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-048-164A-1
Sequence 1, Application US/08048164A
Patent No. 5496934
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shpiegl, Itai
APPLICANT: Goldstein, Marc A.
APPLICANT: Doi, Roy H.
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
City: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/048,164A
FILING DATE: 14-Apr-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..486
US-08-048-164A-1
Query Match 35.2%; Score 270.2; DB 1; Length 486;
Best Local Similarity 98.9%; Pred. No. 3.8e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATGTCAGTTGAATTTCTACACCTTACAAATCAGACACAAACAACTCAATTACACCAATA 60
|||||
DB 16 ATGTCAGTTGAATTTCTACACCTTACAAATCAGACACAAACAACTCAATTACACCAATA 75
OY 61 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTTAGATAT 120
|||||
DB 76 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTTAGATAT 135
OY 121 TATTACACAAAGTATGATGATACACAGCAAACTTCTGCTGTGACATGCTGTGATTA 180
|||||
DB 136 TATTACACAAAGTATGATGATACACAGCAAACTTCTGCTGTGACATGCTGTGATTA 195
OY 181 TTAGGAATTTAGTATGTTGATTAACACTAGCAAAAGTGACACAACTTCTGTTAAAGAAACA 240
|||||
DB 196 TTAGGAATTTAGTATGTTGATTAACACTAGCAAAAGTGACACAACTTCTGTTAAAGAAACA 255
OY 241 GCAAGCCCAACATCACTATGATACATATCTGCA 275
DB 256 GCAAGCCCAACATCACTATGATATATGTTGA 290

RESULT 2

US-08-048-164A-3/C
; Sequence 3, Application US/08048164A
; Patent No. 5486934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Dol, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; US-08-048-164A-3

Query Match 35.2%; Score 270.2; DB 1; Length 486;
Best Local Similarity 98.9%; Pred. No. 3,8e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATGTCAGTTGAATTTCTACACCTTACAAATCAGACACAAACAACTCAATTACACCAATA 60
|||||
DB 471 ATGTCAGTTGAATTTCTACACCTTACAAATCAGACACAAACAACTCAATTACACCAATA 60

OY 61 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTTAGATAT 120
|||||
DB 411 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTTAGATAT 352
OY 121 TATTACACAAAGTATGATGATACACAGCAAACTTCTGCTGTGACATGCTGTGATTA 180
|||||
DB 351 TATTACACAAAGTATGATGATACACAGCAAACTTCTGCTGTGACATGCTGTGATTA 292
OY 181 TTAGGAATTTAGTATGTTGATTAACACTAGCAAAAGTGACACAACTTCTGTTAAAGAAACA 240
|||||
DB 291 TTAGGAATTTAGTATGTTGATTAACACTAGCAAAAGTGACACAACTTCTGTTAAAGAAACA 232
OY 241 GCAAGCCCAACATCACTATGATACATATCTGCA 275
DB 231 GCAAGCCCAACATCACTATGATATATGTTGA 197

RESULT 3

US-08-460-462-1
; Sequence 1, Application US/08460462
; Patent No. 5670623
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Dol, Roy H.
; TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,462
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
; US-08-460-462-1

Query Match 35.2%; Score 270.2; DB 1; Length 486;
Best Local Similarity 98.9%; Pred. No. 3,8e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATGTCAGTTGAATTTCTACACCTTACAAATCAGACACAAACAACTCAATTACACCAATA 60

Db 16 ATGCACTGAATTTTACAACTTAACAATAACAGCAACAACCAATCAATTACACCAATA 75
Qy 61 ATCAAAATTTACTACACATCTGCAGAGATTTTAAATTTAAATGCGTAAAGTTAGATAT 120
Db 76 ATCAAAATTTACTACACATCTGCAGAGATTTTAAATTTAAATGCGTAAAGTTAGATAT 135
Qy 121 TATTACACAGTGTGTGTACACAAGCAAACTTCTGTGTGACCATGCTGGTGCATTA 180
Db 136 TATTACACAGTGTGTGTACACAAGCAAACTTCTGTGTGACCATGCTGGTGCATTA 195
Qy 181 TTAGAAATAGCTATGTTGTAACACTAGCAAACTTCTGTTAAAGAAACA 240
Db 196 TTAGAAATAGCTATGTTGTAACACTAGCAAACTTCTGTTAAAGAAACA 255
Qy 241 GCAAGCCCAACATCACTATGATACATATCTGCA 275
Db 256 GCAAGCCCAACATCACTATGATATATGTTGCA 290

RESULT 4
US-08-460-462-3/C

; Sequence 3, Application US/08460462
; Patent No. 5670623
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shplegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Dol, Roy H.
; TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,462
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; US-08-460-462-3

Query Match 35.2%; Score 270.2; DB 1: Length 486;
Best Local Similarity 98.9%; Pred. No. 3.8e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 ATGCACTGAATTTTACAACTTAACAATAACAGCAACAACCAATCAATTACACCAATA 60
|||||

Db 471 ATGCACTGAATTTTACAACTTAACAATAACAGCAACAACCAATCAATTACACCAATA 412
Qy 61 ATCAAAATTTACTACACATCTGCAGAGATTTTAAATTTAAATGCGTAAAGTTAGATAT 120
Db 411 ATCAAAATTTACTACACATCTGCAGAGATTTTAAATTTAAATGCGTAAAGTTAGATAT 352
Qy 121 TATTACACAGTGTGTGTACACAAGCAAACTTCTGTGTGACCATGCTGGTGCATTA 180
Db 351 TATTACACAGTGTGTGTACACAAGCAAACTTCTGTGTGACCATGCTGGTGCATTA 292
Qy 181 TTAGAAATAGCTATGTTGTAACACTAGCAAACTTCTGTTAAAGAAACA 240
Db 291 TTAGAAATAGCTATGTTGTAACACTAGCAAACTTCTGTTAAAGAAACA 232
Qy 241 GCAAGCCCAACATCACTATGATACATATCTGCA 275
Db 231 GCAAGCCCAACATCACTATGATATATGTTGCA 197

RESULT 5
US-08-460-457-1

; Sequence 1, Application US/08460457
; Patent No. 5719044
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shplegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Dol, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN FUSION PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,457
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
; US-08-460-457-1

Query Match 35.2%; Score 270.2; DB 1: Length 486;
Best Local Similarity 98.9%; Pred. No. 3.8e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 ATGTCAGTTGATTTCTACAACTCTACAAATCAGCAACAACTCAATTACACATA 60
   |||||||
Db 16 ATGTCAGTTGATTTCTACAACTCTACAAATCAGCAACAACTCAATTACACATA 75
QY 61 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAGATAT 120
   |||||||
Db 76 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAGATAT 135
QY 121 TATTACAAAGTGATGATGACACAAAGCAAACTTTCTGCTGATGACATTA 180
   |||||||
Db 136 TATTACAAAGTGATGATGACACAAAGCAAACTTTCTGCTGATGACATTA 195
QY 181 TTAGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
   |||||||
Db 196 TTAGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 255
QY 241 GCAAGCCCAACATCAACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 275
   |||||||
Db 256 GCAAGCCCAACATCAACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 290

```

```

RESULT 6
US-08-460-457-3/C
; Sequence 3, Application US/08460457
; Patent No. 5719044
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpilel, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Dol, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN FUSION PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,457
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEFAX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; US-08-460-457-3

```

```

Query Match      35.2%; Score 270.2; DB 1; Length 486;
Best Local Similarity 98.9%; Pred. No. 3.8e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGATTTCTACAACTCTACAAATCAGCAACAACTCAATTACACATA 60
   |||||||

```

```

Db 471 ATGTCAGTTGATTTCTACAACTCTACAAATCAGCAACAACTCAATTACACATA 412
QY 61 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAGATAT 120
   |||||||
Db 411 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAGATAT 352
QY 121 TATTACAAAGTGATGATGACACAAAGCAAACTTTCTGCTGATGACATTA 180
   |||||||
Db 351 TATTACAAAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 292
QY 181 TTAGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
   |||||||
Db 291 TTAGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 232
QY 241 GCAAGCCCAACATCAACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 275
   |||||||
Db 231 GCAAGCCCAACATCAACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 197

```

```

RESULT 7
US-08-460-458-1
; Sequence 1, Application US/08460458
; Patent No. 5738984
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpilel, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Dol, Roy H.
; TITLE OF INVENTION: KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,458
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEFAX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
; US-08-460-458-1

```

```

Query Match      35.2%; Score 270.2; DB 1; Length 486;
Best Local Similarity 98.9%; Pred. No. 3.8e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGATTTCTACAACTCTACAAATCAGCAACAACTCAATTACACATA 60
   |||||||

```

Db 16 ATGTCACTGTAATTTTACAACTCTAACAAATCAGCAAAACAACTCAATTACACCAATA 75
QY 61 ATCAAAATTTACTACACATCTGACAGTGAATTTAAATTAATGACGTAAGAGTAT 120
Db 76 ATCAAAATTTACTACACATCTGACAGTGAATTTAAATTAATGACGTAAGAGTAT 135
QY 121 TATACACAAGTGTGTGACACAGCAAACTTGTGTGACCAATGCGTGACATTA 180
Db 136 TATACACAAGTGTGTGACACAGCAAACTTGTGTGACCAATGCGTGACATTA 195
QY 181 TTAGGAATAGCTATGTTGATACACTAGCAAGCAAACTTGTTAAAGAAACA 240
Db 196 TTAGGAATAGCTATGTTGATACACTAGCAAGCAAACTTGTTAAAGAAACA 255
QY 241 GCAAGCCCAACATCACTATGATACATATCTGA 275
Db 256 GCAAGCCCAACATCACTATGATACATATCTGA 290

RESULT 8
US-08-460-458-3/C
: Sequence 3, Application US/08460458
: Patent No. 5738984
: GENERAL INFORMATION:
: APPLICANT: Shoseyov, Oded
: TITLE OF INVENTION: KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAIN
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/460,458
: FILING DATE: concurrently herewith
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/048,164
: FILING DATE: 14-APR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Mistrock, S. Leslie
: REGISTRATION NUMBER: 18,872
: REFERENCE/DOCKET NUMBER: 7809-007
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 486 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA
: US-08-460-458-3

Query Match 35.2%; Score 270.2; DB 1; Length 486;
Best Local Similarity 98.9%; Pred. No. 3.8e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCACTGTAATTTTCAAACTCTAACAAATCAGCAAAACAACTCAATTACACCAATA 60
Db 471 ATGTCACTGTAATTTTCAAACTCTAACAAATCAGCAAAACAACTCAATTACACCAATA 412
QY 61 ATCAAAATTTACTACACATCTGACAGTGAATTTAAATTAATGACGTAAGAGTAT 120
|||||

Db 411 ATCAAAATTTACTACACATCTGACAGTGAATTTAAATTAATGACGTAAGAGTAT 352
QY 121 TATACACAAGTGTGTGACACAGCAAACTTGTGTGACCAATGCGTGACATTA 180
Db 351 TATACACAAGTGTGTGACACAGCAAACTTGTGTGACCAATGCGTGACATTA 292
QY 181 TTAGGAATAGCTATGTTGATACACTAGCAAGCAAACTTGTTAAAGAAACA 240
Db 291 TTAGGAATAGCTATGTTGATACACTAGCAAGCAAACTTGTTAAAGAAACA 232
QY 241 GCAAGCCCAACATCACTATGATACATATCTGA 275
Db 231 GCAAGCCCAACATCACTATGATACATATCTGA 197

RESULT 9
US-08-460-455-1
: Sequence 1, Application US/08460455
: Patent No. 5837814
: GENERAL INFORMATION:
: APPLICANT: Shoseyov, Oded
: APPLICANT: Shpiegel, Itai
: APPLICANT: Goldstein, Marc A.
: APPLICANT: Dol, Roy H.
: TITLE OF INVENTION: CELLULOSE BINDING DOMAIN PROTEINS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/460,455
: FILING DATE: concurrently herewith
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/048,164
: FILING DATE: 14-APR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Mistrock, S. Leslie
: REGISTRATION NUMBER: 18,872
: REFERENCE/DOCKET NUMBER: 7809-009
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 486 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..486
: US-08-460-455-1

Query Match 35.2%; Score 270.2; DB 2; Length 486;
Best Local Similarity 98.9%; Pred. No. 3.8e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCACTGTAATTTTCAAACTCTAACAAATCAGCAAAACAACTCAATTACACCAATA 60
Db 16 ATGTCACTGTAATTTTCAAACTCTAACAAATCAGCAAAACAACTCAATTACACCAATA 75
|||||

Oy	61	ATCAAAATTACTAACACATCGACAGGATTTTAAATTAAATGACGAAAGTTACTATN	120
Db	76	ATCAAAATTCTTAACACATCTGCAGAGGATTTAAATTAAATGACGTAAGTTAGATN	135
Oy	121	TATTACACAAAGTGATGGGTACACAAAGAACCTTTCTGCTGTGACCATCGCTGGTGATTA	180
Db	136	TATTACACAAAGTGATGGGTACACAAAGAACCTTTCTGCTGTGACCATCGTGGTGATTA	195
Oy	181	TTAGGAANTAGCTATGTTGATTAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA	240
Db	196	TTAGGAANTAGCTATGTTGATTAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA	255
Oy	241	GCAAGCCCAACATCAACCTTATGATACATATCTGSA	275
Db	256	GCAAGCCCAACATCAACCTTATGATACATATGTTTGA	290

```

US-08-330-394A-3/C
: Sequence 3, Application US/08330394A
: Patent No. 5856201
:
: GENERAL INFORMATION:
: APPLICANT: Shoseyov, Oded
: APPLICANT: Yosef, Karmey
: APPLICANT: Shplegl, Itai
: APPLICANT: Goldstein, Marc A.
: APPLICANT: Dol, Roy H.
: TITLE OF INVENTION: METHODS OF DETECTION USING THE
: TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
: NUMBER OF SEQUENCES: 33
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 115 AVENUE OF THE AMERICAS
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: U.S.A.
: ZIP: 10036
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/330.394A
: FILING DATE: 27-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: CIP OF PCT/US94/04132
: FILING DATE: 14-APR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: MISROCK, S. LESLIE
: REFERENCE/DOCKET NUMBER: 7809-005
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864
: TELEX: 66441 PENNIE
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 486 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA
:
US-08-330-394A-3

```

Oy	61	ATAAAATTCTAACACACTCGACAGGATTTAAATTAAATATGACGTAAGGTTAGCTAT	120
Db	411	ATCAAAATTCTAACACACTCGACAGGATTTAAATTAAATGACGTAAGGTTAGCTAT	352
Oy	121	TATTCACACAGTGAATGGTATACAGAGCAAACTTCTGGTGTGACCATGCTGGTGCAATTA	180
Db	351	TATTCACACAGTGAATGGTATACAGAGCAAACTTCTGGTGTGACCATGCTGGTGCAATTA	292
Oy	131	TTTAGCAAAATACCTATGTTGATTAACACTTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA	240
Db	291	TTTAGCAAAATACCTATGTTGATTAACACTTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA	232
Oy	241	GCAAGCCCCATCATCAACCTATGATACATATTCGTA	275
Db	231	GCAAGCCCCATCATCAACCTATGATACATATTCGTA	197

```

RESULT 13
US-09-006-636-5
: Sequence 5, Application US/09006636
: Patent No. 6005092
: GENERAL INFORMATION:
: APPLICANT: Shoseyov, Oded
: APPLICANT: Shani, Ziv
: TITLE OF INVENTION: ARABIDOPSIS THALIANA ENDO-1,4-BETA-
: TITLE OF INVENTION: GLUCANASE GENE, PROMOTER AND PROTEIN
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennie & Edmonds LLP
: STREET: 115 Avenue of the Americas
: CITY: New York
: STATE: NY
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/006,636
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Baldwin, Geraldine F.
: REGISTRATION NUMBER: 31,232
: REFERENCE/DOCKET NUMBER: 7809-019
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 499 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
:
US-09-006-636-5
Query Match 35.2%; Score 270.2; DB 3; Length 499;
Best Local Similarity 98.9%; Pred. No. 3,9e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAAGTTGAATTCTACACGCTTAACAATAATGACGACAACAACAACTCAATTACACCAATA 60
Db 21 ATGTCAAGTTGAATTCTACACGCTTAACAATAATGACGACAACAACAACCAATTTACACCAATA 80
QY 61 ATCAAAATATACGACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAATGATAT 120
Db 81 ATCAAAATATACGACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAATGATAT 140
QY 121 TATTACACAAGTGATGCTACACAGACACAACCTTCTGTGTGACCAACCTGCTGATTA 180

```

Db 141 TATTACAGAGTGTATGATACAGAGCAAACTTCTGCTGTGACCATGCTGTGATTA 200
OY 181 TTAGGAATATGCTATGCTATGATACACTGCAAGTGAAGCAAACTTCTGTTAAGAAACA 240
Db 201 TTAGGAATATGCTATGCTATGATACACTGCAAGTGAAGCAAACTTCTGTTAAGAAACA 260
OY 241 GCAAGCCCAACATCAACCTATGATACATATGTTGA 275
Db 261 GCAAGCCCAACATCAACCTATGATACATATGTTGA 295

RESULT 14

US-09-006-632-5
Sequence 5, Application US/09006632
Patent No. 6184440
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shani, Ziv
APPLICANT: Shpigel, Eral
TITLE OF INVENTION: TRANSGENIC PLANTS OF ALTERED
TITLE OF INVENTION: MORPHOLOGY
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,632
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7809-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 499 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-006-632-5

Query Match 35.2%; Score 270.2; DB 4; Length 499;
Best Local Similarity 98.9%; Pred. No. 3.9e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATGTCAGTGAATCTCTACACTGCTAACAATCGCAACAACAACTTAATACACCATTA 60
Db 21 ATGTCAGTGAATCTCTACACTGCTAACAATCGCAACAACAACTTAATACACCATTA 80
OY 61 ATCAAAATTTACTTAACATCTGACAGTGAATTTAAATTAATGACGTAAGTTAGTAT 120
Db 81 ATCAAAATTTACTTAACATCTGACAGTGAATTTAAATTAATGACGTAAGTTAGTAT 140
OY 121 TATTACACAAGTATGCTATACACAAGCAAACTTCTGTGTGACCATGCTGTGATTA 180
Db 141 TATTACACAAGTATGCTATACACAAGCAAACTTCTGTGTGACCATGCTGTGATTA 200
OY 181 TTAGGAATATGCTATGCTATGATACACTGCAAGTGAAGCAAACTTCTGTTAAGAAACA 240

Db 201 TTAGGAATATGCTATGCTATGATACACTGCAAGTGAAGCAAACTTCTGTTAAGAAACA 260
OY 241 GCAAGCCCAACATCAACCTATGATACATATGTTGA 275
Db 261 GCAAGCCCAACATCAACCTATGATACATATGTTGA 295

RESULT 15

US-09-277-716-21
Sequence 21, Application US/09277716A
Patent No. 6232107
GENERAL INFORMATION:
APPLICANT: Bryan, Bruce
APPLICANT: Bryant-Gorygy, Christopher
APPLICANT: PROLUME, LTD.
TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
CURRENT FILING DATE: 1999-03-26
EARLIER APPLICATION NUMBER: 60/102,939
EARLIER FILING DATE: 1998-10-01
EARLIER APPLICATION NUMBER: 60/089,367
EARLIER FILING DATE: 1998-06-15
EARLIER APPLICATION NUMBER: 60/079,624
EARLIER FILING DATE: 1998-03-27
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 21
LENGTH: 1146
TYPE: DNA
ORGANISM: Artificial Sequence: fusion construct
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1146)
FEATURE:
OTHER INFORMATION: DNA encoding a Cellulose Binding Domain (CBD)-Gaussia lucifera
US-09-277-716-21

Query Match 35.2%; Score 270.2; DB 4; Length 1146;
Best Local Similarity 98.9%; Pred. No. 6.2e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATGTCAGTGAATCTCTACACTGCTAACAATCGCAACAACAACTTAATACACCATTA 60
Db 1 ATGTCAGTGAATCTCTACACTGCTAACAATCGCAACAACAACTTAATACACCATTA 60
OY 61 ATCAAAATTTACTTAACATCTGACAGTGAATTTAAATTAATGACGTAAGTTAGTAT 120
Db 61 ATCAAAATTTACTTAACATCTGACAGTGAATTTAAATTAATGACGTAAGTTAGTAT 120
OY 121 TATTACACAAGTATGCTATACACAAGCAAACTTCTGTGTGACCATGCTGTGATTA 180
Db 121 TATTACACAAGTATGCTATACACAAGCAAACTTCTGTGTGACCATGCTGTGATTA 180
OY 181 TTAGGAATATGCTATGCTATGATACACTGCAAGTGAAGCAAACTTCTGTTAAGAAACA 240
Db 181 TTAGGAATATGCTATGCTATGATACACTGCAAGTGAAGCAAACTTCTGTTAAGAAACA 240
OY 241 GCAAGCCCAACATCAACCTATGATACATATGTTGA 275
Db 241 GCAAGCCCAACATCAACCTATGATACATATGTTGA 275

Search completed: October 27, 2002, 20:15:44
Job time : 61.152 secs

PD 17-MAR-2001.
 XX
 PF 17-SEP-1999; 99CA-2281913.
 XX
 PR 17-SEP-1999; 99CA-2281913.
 XX
 PA (KAYW/) KAY W. W.
 PA (BURI/) BURIAN J.
 PA (KUZV/) KUZV M. A.
 XX
 PI Kay WW, Burian J, Kuzk MA;
 XX
 DR WPI: 2001-316844/34.
 DR P-PSDB: AAB81127.
 XX
 PT Method for protecting poikilothermic fish against salmonid rickettsial
 PT septicemia and other rickettsial diseases comprises administering a
 PT vaccine containing the Ospa protein of *Piscirickettsia salmonis*.
 XX
 PS Example 4; Fig 5; 35pp; English.
 XX
 CC This invention relates to a method for the protection against infection
 CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*
 CC salmonis. The method comprises administering an immunogenic amount of a
 CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
 CC Ospa in the form of a vaccine. The method is used for protecting animals,
 CC particularly poikilothermic fish, against the bacterial pathogen
 CC P. salmonis. The method is also useful for protecting against salmonid
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
 CC sequence represents P. salmonis Ospa DNA termed C17E2 optimised for
 CC expression in *Escherichia coli* fused to DNA encoding an undefined
 CC N-terminal fusion partner. The protein encoded by this fusion construct
 CC is used in a vaccine to create an anti-Ospa antibody response.
 XX
 SQ Sequence 768 BP; 220 A; 183 C; 190 G; 175 T; 0 other:
 Query Match 100.0%; Score 768; DB 22; Length 768;
 Best Local Similarity 100.0%; Pred. No. 1e-214;
 Matches 768; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1 ATGTGAGTTGATTTCTACAACTTACAAATCAGACAAACAACTCAATACACCAATA 60
 DB 1 ATGTGAGTTGATTTCTACAACTTACAAATCAGACAAACAACTCAATACACCAATA 60
 QY 61 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120
 DB 61 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120
 QY 121 TATTACACAAGTGTGATGACACAAGCAAACTTCTGGTGACCATGCTGTCGATTA 180
 DB 121 TATTACACAAGTGTGATGACACAAGCAAACTTCTGGTGACCATGCTGTCGATTA 180
 QY 181 TTAGGAATAGCTATGTTGATTAACACTAGCAAAAGTGACACAACTTCGTTAAAGAAACA 240
 DB 181 TTAGGAATAGCTATGTTGATTAACACTAGCAAAAGTGACACAACTTCGTTAAAGAAACA 240
 QY 241 GCAAGCCCAATCAACTATGATATCTGATCTGATCCGTCATATGCTGTTGCCCTG 300
 DB 241 GCAAGCCCAATCAACTATGATATCTGATCTGATCCGTCATATGCTGTTGCCCTG 300
 QY 301 CAGGGCAGCTCTCTGATCATTTATCTGTTTCTGTTGGTGGCCGCCGCAAGACTCAGC 360
 DB 301 CAGGGCAGCTCTCTGATCATTTATCTGTTTCTGTTGGTGGCCGCCGCAAGACTCAGC 360
 QY 361 CGCCAGAGAGTTGGCGCGCCACCGGTGCGGTGTTGGCGGTTTCCGCGCAGCTGTTTC 420
 DB 361 CGCCAGAGAGTTGGCGCGCCACCGGTGCGGTGTTGGCGGTTTCCGCGCAGCTGTTTC 420
 QY 421 GGTAAAGGCTCTGGTGTGTCGATGGCCATGGCGGCTGCTGGCGGCTCTGATTT 480
 DB 421 GGTAAAGGCTCTGGTGTGTCGATGGCCATGGCGGCTGCTGGCGGCTCTGATTT 480
 QY 481 GGGCTTAAATCGGTCAGAGCATGAGACAGAGATTAATCAAACTGAACAGTCTCTG 540

DB 481 GGGCTTAAATCGGTCAGAGCATGAGACAGAGATTAATCAAACTGAACAGTCTCTG 540
 QY 541 GAAAGAGTGAAGCCCGCAGAGTTACTGTTGGCGTAATCCGGACCGGTAACAGCTAC 600
 DB 541 GAAAGAGTGAAGCCCGCAGAGTTACTGTTGGCGTAATCCGGACCGGTAACAGCTAC 600
 QY 601 TCTGTGGACCGGTTTCGACCTACCAAGCTTACAAACAAGAGAAAGCCGTCAGCACTAC 660
 DB 601 TCTGTGGACCGGTTTCGACCTACCAAGCTTACAAACAAGAGAAAGCCGTCAGCACTAC 660
 QY 661 TGCCGCAATTTTCAGCAAAAGCCATGATGCGACGTCAGAAACAGAAATCTAGGACACC 720
 DB 661 TGCCGCAATTTTCAGCAAAAGCCATGATGCGACGTCAGAAACAGAAATCTAGGACACC 720
 QY 721 GCGTGCCCTCAGCCGAGATGGCCGCTGCGAGGTGATTAAGCAACCGAATAA 768
 DB 721 GCGTGCCCTCAGCCGAGATGGCCGCTGCGAGGTGATTAAGCAACCGAATAA 768

RESULT 2
 AAF86247
 ID AAF86247 standard; DNA; 483 BP.
 XX
 AC AAF86247;
 XX
 DT 11-JUL-2001 (first entry)
 XX
 DE DNA sequence of E. coli optimised ospa gene 17E2.
 XX
 KW Poikilothermic fish; *Piscirickettsia salmonis*; rickettsial pathogen;
 KW vaccine; OSPA; salmonid rickettsial septicemia; rickettsial disease;
 KW SMS; 17E2; ds.
 XX
 OS *Piscirickettsia salmonis*.
 OS Synthetic.
 OS
 XX
 FT Key 1.483 Location/Qualifiers
 FT CDS
 FT /tag= a
 FT /partial= "Ospa"
 FT /product= "Genus specific 17kDa antigen, the sequence does
 FT /note= "not include a stop codon"
 XX
 CA2281913-11.
 PD 17-MAR-2001.
 XX
 PF 17-SEP-1999; 99CA-2281913.
 XX
 PR 17-SEP-1999; 99CA-2281913.
 XX
 PA (KAYW/) KAY W. W.
 PA (BURI/) BURIAN J.
 PA (KUZV/) KUZV M. A.
 XX
 PI Kay WW, Burian J, Kuzk MA;
 XX
 DR WPI: 2001-316844/34.
 DR P-PSDB: AAB81127.
 XX
 PT Method for protecting poikilothermic fish against salmonid rickettsial
 PT septicemia and other rickettsial diseases comprises administering a
 PT vaccine containing the Ospa protein of *Piscirickettsia salmonis*.
 XX
 PS Example 3; Fig 4C; 35pp; English.
 XX
 CC This invention relates to a method for the protection against infection
 CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*
 CC salmonis. The method comprises administering an immunogenic amount of a
 CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of
 CC Ospa in the form of a vaccine. The method is used for protecting animals,

CC particularly polkilothermic fish, against the bacterial pathogen
 CC P. salmonis. The method is also useful for protecting against salmonid
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
 CC sequence represents DNA which has been optimised for expression in
 CC Escherichia coli to encode the P. salmonis OspA protein. An OspA protein
 CC with an N-terminal fusion partner is used in a vaccine to create an
 CC anti-OspA antibody response.

SO Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;

Query Match 62.9%; Score 483; DB 22; Length 483;
 Best Local Similarity 100.0%; Pred. No. 2,6e-131;

Matches 483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 286 ATGCGTGGTTCCTGCGAGGCGAGCTCTGATCATTTATCTCTTTTCTCGTGGGTTGC 345
 DB 1 ATGCGGTGGTTCCTGCGAGGCGAGCTCTGATCATTTATCTCTTTTCTCGTGGGTTGC 60
 OY 346 GCCCAGACTTTCAGCCGCGAGAAAGTTGGCGCGCCAGCGGTTGGGCGGTTGTT 405
 DB 61 GCCCAGACTTTCAGCCGCGAGAAAGTTGGCGCGCCAGCGGTTGGGCGGTTGTT 120
 OY 406 GCCGCGAGCTGTTCGTAAGGCTCTGCTGCTGATGCGATGCGCATTCGCGGTTG 465
 DB 121 GCCGCGAGCTGTTCGTAAGGCTCTGCTGCTGATGCGCATTCGCGGTTG 180
 OY 466 CTGGCGGCTCTGATTTGGCTCTTAATTCGGTCAGAGCATGAGCAGAGTAATTAATCAA 525
 DB 181 CTGGCGGCTCTGATTTGGCTCTTAATTCGGTCAGAGCATGAGCAGAGTAATTAATCAA 240
 OY 526 CTGAACCACTCTCTGAAAAAGTGAAGCGCGGCGAGTTACTCTGTTGGCGTAAATCCGGAC 585
 DB 241 CTGAACCACTCTCTGAAAAAGTGAAGCGCGGCGAGTTACTCTGTTGGCGTAAATCCGGAC 300
 OY 586 ACCGCTACAGACTCTCTGTGGAACCGGTTGCGACCTACAGCGGTTACAAACAAGAGAA 645
 DB 301 ACCGCTACAGACTCTCTGTGGAACCGGTTGCGACCTACAGCGGTTACAAACAAGAGAA 360
 OY 646 CCGCGTACAGACTCTCTGTGGAACCGGTTGCGACCTACAGCGGTTACAAACAAGAGAA 705
 DB 361 CCGCGTACAGACTCTCTGTGGAACCGGTTGCGACCTACAGCGGTTACAAACAAGAGAA 420
 OY 706 GAAATCTACGCGACCGCGTGCCTCAGCGGATGCGCGTGGCAGGATTAACACCGAA 765
 DB 421 GAAATCTACGCGACCGCGTGCCTCAGCGGATGCGCGTGGCAGGATTAACACCGAA 480
 OY 766 AAA 768
 DB 481 AAA 483

RESULT 3
 AAF86246
 ID AAF86246 standard; DNA; 486 BP.

XX AAF86246;
 AC 11-JUL-2001 (first entry)
 XX
 DE OspA 17kD antigen gene.
 XX
 KW Polkilothermic fish: Piscirickettsia salmonis; rickettsial pathogen;
 KW vaccine; OspA; salmonid rickettsial septicemia; rickettsial disease;
 KW SRS; ds.
 XX
 OS Piscirickettsia salmonis.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..486
 FT /tag= a
 FT /partial
 FT /product= "OspA"
 FT /note= "Genus specific 17kDa antigen, the sequence does

FT not include a stop codon"

XX CA2281913-A1.

XX 17-MAR-2001.

XX 17-SEP-1999; 99CA-2281913.

XX 17-SEP-1999; 99CA-2281913.

XX (KAYW/) KAY W W.

XX (BURI/) BURIAN J.

XX (KUZK/) KUZK M A.

XX Kay WW, Burian J, Kuzk MA;

XX WPI: 2001-316844/34.

XX P-PSDB: AAB81126.

PT Method for protecting polkilothermic fish against salmonid rickettsial
 PT septicemia and other rickettsial diseases comprises administering a
 PT vaccine containing the OspA protein of Piscirickettsia salmonis

Disclosure; Fig 28; 35pp; English.

CC This invention relates to a method for the protection against infection
 CC of a polkilothermic fish by the bacterial pathogen, Piscirickettsia
 CC salmonis. The method comprises administering an immunogenic amount of a
 CC P. salmonis specific antigen termed OspA, or an immunogenic fragment of
 CC OspA in the form of a vaccine. The method is used for protecting animals,
 CC particularly polkilothermic fish, against the bacterial pathogen
 CC P. salmonis. The method is also useful for protecting against salmonid
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
 CC sequence represents DNA encoding the P. salmonis OspA protein. An OspA
 CC protein with an N-terminal fusion partner is used in a vaccine to create
 CC an anti-OspA antibody response.

SO Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other;

Query Match 38.4%; Score 295; DB 22; Length 486;
 Best Local Similarity 76.0%; Pred. No. 3.1e-76;

Matches 364; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

OY 290 GTGGTTCCTGCAAGGCGAGCTCTGATCATTTATCTCTGTTCCGTGGGTTGCGCC 349
 DB 8 GAGGATCTTTCAGAGGAGTATGCTATTAATTAATCACTGTTTAACTGCTGCGCC 67
 OY 350 AGAAGTTTCAGCCGCCAGAGAGTTGGCGCGCCAGCGTGGTGGGCGGTTGCGCC 409
 DB 68 AGAAGTTTCAGCCGCCAGAGAGTTGGCGCGCCAGCGTGGTGGGCGGTTGCGCC 127
 OY 410 GCCAGCTGTTCGTAAGGCTCTGTCGTGTCGATGCGCATTCGCGGTTGCGG 469
 DB 128 GCCAGCTGTTCGTAAGGCTCTGTCGTGTCGATGCGCATTCGCGGTTGCGG 187
 OY 470 GCGGTCGATTTGCTCAAAATGCGTCAAGAGCATTCGAGCAGAGGATTAATCAACTGA 529
 DB 188 GTGGATTAATTTGCTCAAAATGCGTCAAGAGCATTCGAGCAGAGGATTAATCAACTGA 247
 OY 530 ACCAGTCTTGAAAAAGTGAAGCGCGCCAGGTTACTGCTGGCGTAATCCGAGACCG 589
 DB 248 ACCAGTCTTGAAAAAGTGAAGCGCGCCAGGTTACTGCTGGCGTAATCCGAGATTCAG 307
 OY 590 GTAACAGACTACTCTGTGGAACCGGTTGCGACCTACAGCGTTACAAACAAGAGACCG 649
 DB 308 GCAATAGTATTACTGTGAGACCGAGTCTACTTACAGCGTTACAAACAAGAGACGTC 367
 OY 650 GTACAGACTCTGCGGATTCAGAGAAAGCATTCGAGGATTCAGAGGATTCAGAGGATTCAG 709
 DB 368 GCCAGCAATATTTGTCAGGAATTTTCAGCAAAAGCGATGATTCAGAGGATTCAGAGGATTCAG 427
 OY 710 TCTACGACACCGGCTCCCTCAGCCGAGAGGCGGTCGAGGATTCAGGATTCAGGATTCAG 768

Db	188	GTGATTAAATTTGGTTCTAAATAATCGGTCATCGATGTGATGACACAGATAAATAATAAAGCTAA	247
Oy	530	ACCAAGTCTCGAAAAAGTAAAGCCGCGCAGAGTTACTGCTTGGCCGTAAATCCGACACCG	589
Db	248	ACCAAGTATTGGAAAGAGTAAAGCAAGGCAAGTGACACGTTGGCGTAATCCAGATACAG	307
Oy	590	GTAACAGCTCTCTGTGTGAGAACCGGTTTCGACCCTACACAGCGTTACAAACAAACAGAAACGCC	649
Db	308	GCAATAGTGTAACTGTGTGGCCAGTGCGCTACTTACACAGCGTTACAAATTAAGCAAGACGCTC	367
Oy	650	GTCAGCAGTACTCCGCCGATTTACAGCAAGAACGCATGTCGAGGTCAGAAACAGGAA	709
Db	368	GCCAGCAAAATATGTGAGATATTTACAGAAAGGGGATGATTGCAGGGCAGAACAGACAGA	427
Oy	710	TCACAGGCAACCGGTCGCTTACGCGGATGTGCGCGCTGCGACAGTGATTTACACCGAA	768
Db	428	TTTACGGCACTGCATGCGCGCAACCGGATGTGCTGTGCAAGTCATTTCAACAGAAAA	486
RESULT 5			
AAD11043			
ID	AAD11043	standard; DNA; 573 BP.	
XX			
AC	AAD11043;		
XX			
DT	24-SEP-2001	(first entry)	
XX			
DE	Clostridium cellulovorans	cellulose binding domain-180 (CBD-180)	DNA.
XX			
M	Polysaccharide modification; polysaccharide binding domain; PBD; paper		
KW	yarn, fiber; textiles; biological crosslinker; mechanical property;		
KW	Clostridium cellulovorans cellulose binding domain; wet strength;		
KM	durability; elasticity; CBDc1os; cellulose binding protein A; CBP A;		
XX	CBD-180; ds.		
XX			
OS	Clostridium cellulovorans.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	3..572	
FT		/*tag= a	
FT		/product= "Cellulose binding domain-180 protein"	
FT		/note= "CDS does not include stop codon"	
XX		/partial	
PN	MO200134091-A2.		
XX			
PD	17-MAY-2001.		
XX			
PF	02-NOV-2000; 2000MO-IL00708.		
XX			
PR	08-NOV-1999; 99US-0164140.		
PR	18-NOV-1999; 99US-0166389.		
XX			
PA	(CBDT-) CBD TECHNOLOGIES LTD.		
PA	(T1SS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.		
PI	Levy I, Nussinovitch A, Shoseyov O;		
XX			
DR	WPI: 2001-457121/49.		
XX	P-PSDB; AAE05746.		
PT	Preparation of a polysaccharide containing material having at least one		
PT	desired structural, chemical, physical, electrical and/or mechanical		
XX	property		
XX			
PS	Example 1.2; Fig 1e-1g; 121pp; English.		

CC mechanical properties of polysaccharide materials such as paper,
CC yarns, fibers and textiles, using biological crosslinking agents.
CC The polysaccharide containing materials have improved mechanical
CC properties such as wet strengths, durability and elasticity. The PBD
CC reagent is applied in the forming stage in fluting paper manufacture
CC which eliminates the sizing step. The use of a biological crosslinker
CC improves the recyclability of paper products. The PBD reagent maintains
CC the fine fibers in a slurry therefore resulting in better recovery of
CC raw materials. The PBD molecules are eluted by strong alkaline conditions
CC which enhances the ability of the alkaline glue used in binding to
CC penetrate paper.
CC The present sequence is a DNA encoding Clostridium cellulovorans
CC cellulose binding domain (CBD)-180. CBD-180 contains 180 amino acids of
CC Clostridium cellulovorans CBD (CBDClos) of cellulose binding protein A
CC (CBP A).
CC
XX Sequence 573 BP; 215 A; 112 C; 98 G; 148 T; 0 other;
SQ
Query Match 35.4%; Score 271.8; DB 22; Length 573;
Best Local Similarity 99.3%; Pred. No. 2.1e-69;
Matches 273; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
0;
QY 1 ATGTCAGTGAATTCACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 60
DB 3 ATGTCAGTGAATTCACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 62
QY 61 ATCAAAATTTACTACACATCTGACAGATTTAAATTAATGCGTAAAGTAGATAT 120
DB 63 ATCAAAATTTACTACACATCTGACAGATTTAAATTAATGCGTAAAGTAGATAT 122
QY 121 TATTACACAAGTGTGTACAGAGCAAACTTCTGCTGTGACCATTCGTGTCATTA 180
DB 123 TATTACACAAGTGTGTACAGAGCAAACTTCTGCTGTGACCATTCGTGTCATTA 182
QY 181 TTAGGAATAGCTATGTTGATACACTAGCAAGAGCAAACTTCGTTAAAGAAACA 240
DB 183 TTAGGAATAGCTATGTTGATACACTAGCAAGAGCAAACTTCGTTAAAGAAACA 242
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGCA 275
DB 243 GCAAGCCCAACATCAACCTATGATACATATCTGCA 277
RESULT 6
AAD11044
ID AAD11044 standard; DNA: 1030 BP.
XX
AC AAD11044:
XX
DT 24-SEP-2001 (first entry)
XX
DE Clostridium cellulovorans CBD cross linker protein (CCP) encoding DNA.
XX
DE
XX
KM Polysaccharide modification; polysaccharide binding domain; PBD; paper;
KM yarn; fiber; textile; biological crosslinker; mechanical property;
KM wet strength; durability; elasticity; cellulose binding domain; CBD;
KM CBD cross linker protein; CCP; ds.
XX
OS Clostridium cellulovorans.
XX
FH Key location/Qualifiers
FT CDS 3..1028
FT
FT /*tag= a
FT /product= "CCP protein"
FT /transl_except= "(pos: 1020..1028, aa:Pro-Asp)"
FT /note= "CDS does not include stop codon"
FT /partial
XX
XX
PN MO200134091-A2.
XX
XX 17-MAY-2001.
XX
PF 02-NOV-2000; 2000MO-IL00708.

XX
PR 08-NOV-1999; 99US-0164140.
PR 18-NOV-1999; 99US-0166389.
XX
XX
PA (CBDT-) CBD TECHNOLOGIES LTD.
PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
PI Levy I, Nussimovitch A, Shoseyov O;
XX
XX WPI; 2001-457121/49.
DR P-PSDB; AAE05747.
XX
XX Preparation of a polysaccharide containing material having at least one
PT desired structural, chemical, physical, electrical and/or mechanical
PT property -
XX
XX Example 1.2; Fig 2b-2e; 121pp; English.
XX
XX The present invention relates to methods and compositions for cross-
CC linking and/or modifying the properties of polysaccharide materials.
CC The method involves treating the polysaccharide structure with a
CC polysaccharide binding domain (PBD) fusion protein. The method is
CC used to alter the structural, chemical, physical, electrical and
CC mechanical properties of polysaccharide materials such as paper,
CC yarns, fibers and textiles, using biological crosslinking agents.
CC The polysaccharide containing materials have improved mechanical
CC properties such as wet strengths, durability and elasticity. The PBD
CC reagent is applied in the forming stage in fluting paper manufacture
CC which eliminates the sizing step. The use of a biological crosslinker
CC improves the recyclability of paper products. The PBD reagent maintains
CC the fine fibers in a slurry therefore resulting in better recovery of
CC raw materials. The PBD molecules are eluted by strong alkaline conditions
CC which enhances the ability of the alkaline glue used in binding to
CC penetrate paper.
CC The present sequence is a DNA encoding Clostridium cellulovorans
CC cellulose binding domain (CBD) cross linker protein (CCP).
CC
XX
SQ Sequence 1030 BP; 402 A; 195 C; 166 G; 267 T; 0 other;
Query Match 35.4%; Score 271.8; DB 22; Length 1030;
Best Local Similarity 99.3%; Pred. No. 2.9e-69;
Matches 273; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
0;
QY 1 ATGTCAGTGAATTCACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 60
DB 3 ATGTCAGTGAATTCACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATA 62
QY 61 ATCAAAATTTACTACACATCTGACAGATTTAAATTAATGAGCTAAAGTAGATAT 120
DB 63 ATCAAAATTTACTACACATCTGACAGATTTAAATTAATGAGCTAAAGTAGATAT 122
QY 121 TATTACACAAGTGTGTACAGAGCAAACTTCTGCTGTGACCATTCGTGTCATTA 180
DB 123 TATTACACAAGTGTGTACAGAGCAAACTTCTGCTGTGACCATTCGTGTCATTA 182
QY 181 TTAGGAATAGCTATGTTGATACACTAGCAAGAGCAAACTTCGTTAAAGAAACA 240
DB 183 TTAGGAATAGCTATGTTGATACACTAGCAAGAGCAAACTTCGTTAAAGAAACA 242
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGCA 275
DB 243 GCAAGCCCAACATCAACCTATGATACATATCTGCA 277
RESULT 7
AAQ72917
ID AAQ72917 standard; DNA: 486 BP.
XX
AC AAQ72917:
XX
DT 05-JUN-1995 (first entry)
XX
XX Cellulose binding domain.
DE

```

XX XX Cellulose binding domain; CBD; ds.
XX KM Clostridium cellulovorans.
XX OS
XX XX
XX XX
XX FH Key Location/Qualifiers
XX FT 1..486
XX FT CDS /*tag= a
XX
XX PN MO9424158-A.
XX
XX XX
XX XX 27-OCT-1994.
XX
XX XX
XX PF 14-APR-1994; 94MO-US04132.
XX
XX PR 14-APR-1993; 93US-0048164.
XX
XX PA (REGC ) UNIV CALIFORNIA.
XX PA (YISS ) YISSUM RES & DEV CO.
XX
XX PI Doi RH, Goldstein MA, Shoseyov O, Shpiegl I;
XX
XX DR WPI; 1994-341767/42.
XX DR P-PSDB; AAR62634.
XX
XX PT Isolated cellulose binding domain and fusion proteins - with
XX PT applications. Including drug delivery, affinity separations, and
XX PT diagnostic techniques
XX
XX PS Claim 10; Fig 1; 125pp; English.
XX
XX CC AA072917/R63634 is a novel isolated cellulose binding domain. It pref.
XX CC binds crystalline with a Kd ranging from 1.5 microm - 0.5 microm,
XX CC pref. with a Kd of <1.2 microm, esp. <1.0 microm. The protein and
XX CC the encoding DNA are claimed, as is a CBD fusion protein (FP)
XX CC comprising the CBD and a second protein. The second protein is pref.
XX CC Protein A, heparinase, a hormone or an enzyme capable of degrading
XX CC an environmental pollutant, or an HSP. HSP antibody, cross-reactive
XX CC HSP-related protein or an antigenic portion of this. The CBD and FP
XX CC may be used in drug delivery, affinity seps. and diagnostic
XX CC techniques. CBD nucleic acid may be obtd. from a variety of cell
XX CC sources that produce CBDs that bind with high affinity and in a
XX CC reversible manner or that produce CBD encoding mRNA. The preferred
XX CC source of CBD encoding nucleic acid is C. cellulovorans.
XX
XX SQ Sequence 486 BP; 194 A; 89 C; 79 G; 124 T; 0 other;

Query Match 35.2%; Score 270.2; DB 15; Length 486;
Best Local Similarity 98.9%; Pred. No. 5,7e-69;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATGTGAGTTGAATTTCTACACTTAACAATCAGCACAACAACCTCAATTACCCATA 60
DB 16 ATGTGAGTTGAATTTCTACACTTAACAATCAGCACAACAACCTCAATTACCCATA 75
OY 61 ATCAAAATTTACTAACACATCTGCAGTGATTTAAATTAATGACGTAAAGTTAGATAT 120
DB 76 ATCAAAATTTACTAACACATCTGCAGTGATTTAAATTAATGACGTAAAGTTAGATAT 135
OY 121 TATTACACAAGTGTGATGACAGCAAGCAAACTTCTGCTGTCACCAATGCTGTGCATTA 180
DB 136 TATTACACAAGTGTGATGACAGCAAGCAAACTTCTGCTGTCACCAATGCTGTGCATTA 195
OY 181 TTAGGAATAGCTATGTTGATTAACACAGCAAGTGCACCAAACTTCGTTAAAGAAACA 240
DB 196 TTAGGAATAGCTATGTTGATTAACACAGCAAGTGCACCAAACTTCGTTAAAGAAACA 255
OY 241 GCAAGCCCAATCACTATGATATACATATCTGGA 275
DB 256 GCAAGCCCAATCACTATGATATGATATGTTGA 290

```

RESULT 8

```

AAV74072
ID AAV74072 standard; DNA; 486 BP.
XX
XX AC AAV74072;
XX
XX XX
XX DT 09-MAR-1999 (first entry)
XX
XX DE C. cellulovorans Cbpa CBD DNA.
XX
XX XX
XX KM Cellulose binding domain; CBD; Cbpa; fusion protein; diagnosis; HSP;
XX KM immunosassay; heat-shock protein; cross reactive protein; detection;
XX KM antigenic fragment; antibody; insulin-dependent diabetes mellitus;
XX KM cellulose; chitin; cellulolytic; amorphogenic; ds.
XX
XX OS Clostridium cellulovorans.
XX
XX FH Key Location/Qualifiers
XX FT 1..486
XX FT CDS /*tag= a
XX FT /product= "cellulose binding domain"
XX
XX XX
XX PN US5856201-A.
XX
XX PD 05-JAN-1999.
XX
XX PF 27-OCT-1994; 94US-0330394.
XX
XX PR 27-OCT-1994; 94US-0330394.
XX PR 14-APR-1993; 93US-0048164.
XX PR 14-APR-1994; 94MO-US04132.
XX
XX PA (REGC ) UNIV CALIFORNIA.
XX PA (YISS ) YISSUM RES & DEV CO.
XX
XX PI Doi RH, Goldstein MA, Shoseyov O, Shpiegl I;
XX
XX DR WPI; 1999-105130/09.
XX DR P-PSDB; AAW90077.
XX
XX PT Detection of a specific analyte by reaction with binding agent fused
XX PT to cellulose binding domain - and subsequent treatment with
XX PT cellulose and reaction of insoluble product with a label specific
XX PT for the analyte
XX
XX PS Example 7.2.1; Fig 1A-B; 63pp; English.
XX
XX CC This sequence encodes a cellulose binding domain (CBD) derived from the
XX CC Cbpa protein of Clostridium cellulovorans. The sequence is used the
XX CC construction of a fusion protein which can be used in diagnostic
XX CC immunosassays, e.g. to detect heat-shock proteins (HSP) and their
XX CC cross-reactive proteins, antigenic fragments or HSP-specific antibodies
XX CC (which indicate insulin-dependent diabetes mellitus, or susceptibility
XX CC to it. The CBD, has very high affinity for cellulose (including
XX CC crystalline forms) and chitin (dissociation constant 0.8-1.4 mu M), but
XX CC has almost no cellulolytic or amorphogenic activities. The CBD binds over
XX CC a wide pH range and is not released from cellulose by washing with water.
XX
XX SQ Sequence 486 BP; 194 A; 89 C; 79 G; 124 T; 0 other;

Query Match 35.2%; Score 270.2; DB 20; Length 486;
Best Local Similarity 98.9%; Pred. No. 5,7e-69;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATGTGAGTTGAATTTCTACACTTAACAATCAGCACAACAACCTCAATTACCCATA 60
DB 16 ATGTGAGTTGAATTTCTACACTTAACAATCAGCACAACAACCTCAATTACCCATA 75
OY 61 ATCAAAATTTACTAACACATCTGCAGTGATTTAAATTAATGACGTAAAGTTAGATAT 120
DB 76 ATCAAAATTTACTAACACATCTGCAGTGATTTAAATTAATGACGTAAAGTTAGATAT 135
OY 121 TATTACACAAGTGTGATGACAGCAAGCAAACTTCTGCTGTCACCAATGCTGTGCATTA 180

```


Db 136 TATTACACAAGTGTGATGACACAGCAAACTTCTGCTGTGACCATGCTGGTCATT 195

QY 181 TTAGSAAATAGCTATGTGTATACACTAGCAAGTGCAGCAACTTCGTTAAGAACA 240

Db 196 TTAGGAAATAGCTATGTGTATACACTAGCAAGTGCAGCAAACTTCGTTAAGAACA 255

QY 241 GCAGCCCAACATCAACCTATGATACATATCTGGA 275

Db 256 GCAGCCCAACATCAACCTATGATACATATGTTGA 290

RESULT 9

AAAX24930 standard; DNA; 499 BP.

XX AAAX24930;

AC AAAX24930;

DT 21-JUN-1999 (first entry)

XX

DE Clostridium cellulovorans cbpa cellulose binding domain DNA.

XX

KW Cellulose binding domain; CBD; cbpa; endo-1,4-beta-glucanase; Cell;

KW Arabidopsis thaliana; transgenic plant; crop improvement;

KW morphology; cell wall; ds.

XX

OS Clostridium cellulovorans.

XX

PN WO9907830-A1.

XX

PD 18-FEB-1999.

XX

PF 26-JUL-1998; 98WO-IL00345.

XX

PR 13-JAN-1998; 98US-0006636.

PR 27-JUL-1997; 97IL-0121404.

PR 13-JAN-1998; 98US-0006632.

XX

PA (YISS) YISSUM RES & DEV CO.

XX

PI Shani Z, Shoseyov O, Shplegl E;

DR WPI: 1999-180488/15.

XX

PT Transgenic plants expressing cell-wall modulating protein - have

PT altered morphology, e.g. increased growth, modified fiber length or

PT cellulose content

XX

PS Disclosure; Page 134; 144pp; English.

XX

CC This DNA fragment encodes the cellulose binding domain (CBD) of

CC the cbpa protein of Clostridium cellulovorans. It was obtained

CC by PCR amplification (see also AAAX24932-53), and was used in the

CC construction of binary vector pCCI, in which cbd was joined to

CC the promoter and signal region of the novel endo-1,4-beta-glucanase

CC cell gene (see AAAX24923) of Arabidopsis thaliana. Expression of cbd

CC in transgenic tobacco plants modulated their growth. The cell

CC promoter can provide expression of any protein in elongating

CC tissue. CBD is an example of a cell wall modulation transgene

CC used to alter the structure or morphology of a plant. Transgenic

CC plants of the invention may have altered biomass, growth, yield,

CC greater or less resistance to biodegradation, be more or less

CC digestible by ruminants, have altered cellulose content, larger or

CC smaller leaves, etc., when compared to non-transgenic plants of the

CC same species.

XX

SQ Sequence 499 BP; 197 A; 93 C; 82 G; 127 T; 0 other;

QY Query Match 35.2%; Score 270.2; DB 20; Length 499;

Best Local Similarity 98.9%; Pred. No. 5.8e-69;

Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAAGTTGAATTTTACAACTTACCAAAATCGCAAAACATTCATTACACCAATA 60

|||||

Db 21 ATGTCAAGTTGAATTTTACAACTTACCAAAATCGCAAAACATTCATTACACCAATA 80

QY 61 ATCAAAATTAAGTAACTAAGCATCTGACAGTATTTAAATTAATGCTAAAGTAACTAT 120

Db 81 ATCAAAATTAAGTAACTAAGCATCTGACAGTATTTAAATTAATGCTAAAGTAACTAT 140

QY 121 TATTACACAAGTGTGATGACACAGCAAACTTCTGCTGTGACCATGCTGGTCATT 180

Db 141 TATTACACAAGTGTGATGACACAGCAAACTTCTGCTGTGACCATGCTGGTCATT 200

QY 181 TTAGSAAATAGCTATGTGTATACACTAGCAAGTGCAGCAAACTTCGTTAAGAACA 240

Db 201 TTAGSAAATAGCTATGTGTATACACTAGCAAGTGCAGCAAACTTCGTTAAGAACA 260

QY 241 GCAGCCCAACATCAACCTATGATACATATCTGGA 275

Db 261 GCAGCCCAACATCAACCTATGATACATATGTTGA 295

RESULT 10

AAD11042 standard; DNA; 507 BP.

XX AAD11042;

AC AAD11042;

DT 24-SEP-2001 (first entry)

XX

DE Clostridium cellulovorans cellulose binding domain (CBDclos) DNA.

XX

KW Polysaccharide modification; polysaccharide binding domain; PBD; paper;

KW yarn; fiber; textile; biological crosslinker; mechanical property;

KW Clostridium cellulovorans cellulose binding domain; wet strength;

KW durability; elasticity; CBDclos; cellulose binding protein A; CBP A; ds.

XX

OS Clostridium cellulovorans.

XX

FH Key Location/Qualifiers

FT CDS 3..494

FT /tag= a

FT /product= "CBDclos protein"

XX

PN WO200134091-A2.

XX

PD 17-MAY-2001.

XX

PF 02-NOV-2000; 2000WO-IL00708.

XX

PR 08-NOV-1999; 99US-0164140.

PR 18-NOV-1999; 99US-0166389.

XX

PA (CBDT-) CBD TECHNOLOGIES LTD.

PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.

PI Levy I, Nussinovitch A, Shoseyov O;

XX

DR WPI: 2001-457121/49.

DR P-PDB: AA05745.

XX

PT Preparation of a polysaccharide containing material having at least one

PT desired structural, chemical, physical, electrical and/or mechanical

PT property -

XX

PS Example 1.1; Page 111; 121pp; English.

XX

CC The present invention relates to methods and compositions for cross-

CC linking and/or modifying the properties of polysaccharide materials.

CC The method involves treating the polysaccharide structure with a

CC polysaccharide binding domain (PBD) fusion protein. The method is

CC used to alter the structural, chemical, physical, electrical and

CC mechanical properties of polysaccharide materials such as paper,

CC yarns, fibers and textiles, using biological crosslinking agents.

CC The polysaccharide containing materials have improved mechanical

CC properties such as wet strengths, durability and elasticity. The PBD

Query Match	35.28;	Score 270.2;	DB 22;	Length 507;
Best Local Similarity	98.98;	Pred. No. 5.8e-69;		
Matches 272;	Conservative	0;	Mismatches 3;	Indels 0;
				Gaps 0

QY	1	ATGTCAGGTGAAATTTCACAACTCTAACAAATCAGCACAAACAAACCAATTCACCAATA	60
Db	21	ATTTCAGTGGAAATTTTACCAACTCTAACAAATCAGCACAAACAAACCAATTCACCAATA	80
QY	61	ATCAAAATTAACTAACACATCTGACAGGATTTAAATTTAAATGACGTAAAGTTAGATAT	120
Db	81	ATCAAAATTAACTAACACATCTGACAGGATTTAAATTTAAATGACGTAAAGTTAGATAT	140
QY	121	TATTACACAGAGTATGATGACACAAAGCAAACTTTCTGGTGTACATGCTGGTGCATTA	180
Db	141	TATTACACAGAGTATGATGACACAAAGCAAACTTTCTGGTGTACATGCTGGTGCATTA	200
QY	181	TTAGGAAATAGCTATGTTGGTAACACATAGCAAAAGTACAGCAAACTTGTTAAAGAAACA	240
Db	201	TTAGGAAATAGCTATGTTGGTAACACATAGCAAAAGTACAGCAAACTTGTTAAAGAAACA	260
QY	241	GCAAGCCCAACATCAACCTATGATACATATCTGGA	275
Db	261	GCAAGCCCAACATCAACCTATGATACATATCTGGA	295

Key	Location/Qualifiers
CDS	1..984
key	Location/Qualifiers
	1..984
	/*tag- a
	/product- "Spep-CBD-Sprot protein"
	/transl_except- "(pos: 979..981, aa:Xaa)"
	/note- "Xaa corresponds to in-frame stop codon;"
	CDS does not include start and stop codon"
	/partial
	68..624
	/*tag- b
	/note- "This region is derived from C. cellulovorans"
	652..981
	/*tag- c
	/note- "This region is derived from bovine"

XX 17-MAY-2001.
PD
XX
XX 02-NOV-2000; 2000WO-IL00708.
PF
XX
XX 08-NOV-1999; 99US-0164140.
PR
XX 18-NOV-1999; 99US-0166389.
XX
XX
XX (CBDT-) CBD TECHNOLOGIES LTD.
PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
XX
PI Levy I, Nussinovitch A, Shoseyov O;
XX
XX
XX WPI; 2001-457121/49.
XX
XX P-PSDB; AAE05749.
XX
XX

PT	Preparation of a polysaccharide containing material having at least one
PI	desired structural, chemical, physical, electrical and/or mechanical
PI	property -
XX	
PS	Example 4; Fig 4b-4g; 121pp; English.
XX	

The present invention relates to methods and compositions for cross-linking and/or modifying the properties of polysaccharide materials. The method involves treating the polysaccharide structure with a polysaccharide binding domain (PBD) fusion protein. The method is used to alter the structural, chemical, physical, electrical and mechanical properties of polysaccharide materials such as paper, yarns, fibers and textiles, using biological crosslinking agents. The polysaccharide containing materials have improved mechanical properties such as wet strengths, durability and elasticity. The PBD reagent is applied in the forming stage in futing paper manufacture which eliminates the sizing step. The use of a biological crosslinker improves the recyclability of paper products. The PBD reagent maintains the fine fibers in a slurry therefore resulting in better recovery of raw materials. The PBD molecules are eluted by strong alkaline conditions which enhances the ability of the alkaline glue used in binding to penetrate paper.

The present sequence is a DNA encoding S peptide-cellulose binding domain-S protein (Speg-CBD-Sprot), a fusion protein derived from Clostridium cellulovorans and bovine.

50 Sequence 984 BP; 325 A; 240 C; 193 G; 226 T; 0 other;

Query Match	35.28;	Score 270.2;	DB 22;	Length 984;
Best Local Similarity	98.9%;	Pred. No. 8.3e-69;		
Matches 272; Conservative	0;	Mismatches 3;	Totals 0	

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

RESULT 12	
AAZ27550	
ID	AAZ27550 standard; DNA; 1146 BP
XX	

AC	AA227550;
XX	
DT	13-DEC-1999 (first entry)
XX	
DE	Gaussia luciferase fusion protein coding sequence.
XX	
KW	Luciferase; green fluorescent protein; GFP; screening assay; diagnosis;
KM	bioluminescence-generating system; toy; cosmetic; fairy dust; beverage;
KW	body paint; squirt gun; balloon; slimy play material; soap; toothpaste;
KM	fusion protein; ds.
XX	
OS	Gaussia sp.
XX	
FM	WO9949019-A2.
PD	
XX	
PD	30-SEP-1999.
XX	
PF	26-MAR-1999; 99WO-US06698.
PR	
XX	
PR	27-MAR-1998; 98US-0079624.
PR	15-JUN-1998; 98US-0089367.
XX	
PR	01-OCT-1998; 98US-0102939.
PA	(PROL-) PROLINE LTD.
PA	(BRYA.) BRYAN B J.
XX	
P1	Bryan BJ, Szent-Gyorgyi C;
DR	WI: 1999-580443/49.
DR	P-PDB: AAY39952.
XX	
PT	New isolated Renilla mulleri, Gaussia and Pleuromamma luciferase and
XX	Renilla and Ptilosarcus green fluorescent protein nucleic acids -
PS	Disclosure: Page 222-223; 23pp; English.
XX	
CC	This sequence encodes a luciferase of the invention. The invention
CC	relates to Renilla mulleri, Gaussia and Pleuromamma luciferase and
CC	Renilla and Ptilosarcus green fluorescent protein (GFP) nucleic acids and
CC	proteins. The luciferases and GFPs can be used in
CC	bioluminescence-generating systems, assays, screening methods, diagnostic
CC	method and articles of manufacture. They can be expressed using
CC	e.g. bacterial, yeast, fungal, plant, insect or animal cells. The Renilla
CC	mulleri, Gaussia and Pleuromamma luciferase or Renilla or Ptilosarcus
CC	GFPs can be used in e.g. toys, cosmetics, fountains, personal care items,
CC	fairy dust, beverages, soft drinks, foods, textile products, bubbles,
CC	balloons, personal items, dentifrices, soaps, body paints, bubble bath,
CC	ink or paper products. In particular, they can be used in e.g. squirt
CC	guns, pellet guns, finger paints, foot bags, greeting cards, slimy play
CC	material, clothing, bubble making toys, bath powders, cosmetics, body
CC	lotions, gels, body powders, body creams, toothpastes, mouthwashes,
CC	soaps, body paints, bubble bath,inks, wrapping paper, gelatins, icings,
CC	frostings, greeting cards, beer, wine, champagne, soft drinks, ice cubes,
CC	ice, dry ice or fountains. The nucleic acids can also be used to produce
CC	transgenic fish and plants.
XX	
SQ	Sequence 1146 BP; 396 A; 222 C; 247 G; 281 T; 0 other;
XX	
Query Match	35.2%; Score 270.2; DB 20; Length 1146;
Best Local Similarity	98.9%; Pred. No. 9e-65; 3; Indels 0; Gaps 0
Matches 272:	Conservative 0; Mismatches 3; Indels 0; Gaps 0
OY	1 ATGTAGATTGAATTTTACAACTCTTAACAATGACGCAACAACATCAATTACCAATA 60
DB	1 ATGTGACTTGAAATTTTACAACTCTTAACAATGACGCAACAACATCAATTACCAATA 60
OY	61 ATCAAATAATTTACTAACACATCTGACAGTGAATTTAAATGAGCTAAAAGTTAGATAT 120
DB	61 ATCAAATAATTTACTAACACATCTGACAGTGAATTTAAATGAGCTAAAAGTTAGATAT 120
OY	121 TATTTCACAGAAGTAGTGATCACAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
DB	121 TATTTCACAGAAGTAGTGATCACAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180

QY	181	TTAGCAATACGTATGTTGATACACATGACAAAGTGCACGAAACCTTCGTTAAAGAAACA	24
DB	181	TTTAGCAATACGTATGTTGATACACATGACAAAGTGCACGAAACCTTCGTTAAAGAAACA	240
QY	241	GCAAGCCCAACATCAACCTATGATACATATCTGGA	275
DB	241	GCAAGCCCAACATCAACCTATGATACATATCTGGA	275
RESULT 13			
AAD22201			
ID	AAD22201	standard; DNA; 1146 BP.	
XX			
AC	AAD22201:		
XX			
DT	12-FEB-2002	(first entry)	
XX			
DE	Gaussia species CBD-luciferase fusion protein encoding DNA.		
XX			
KM	Green fluorescent protein; GFP; bioluminescence generating system; toy		
KM	luciferase; finger paint; slimy play material; fishing lure; sparkler		
KM	doll; balloon; personal care item; cosmetic; bath powder; body cream;		
KM	tooth paste; mouth wash; soap; body paint; ornamental transgenic plan		
KM	bubble bath; fountain; fairy dust; food; gelatin; icing; frosting; bea		
KM	wine; champagne; milk; soft drink; ice cube; ice; foot bag; beverage;		
KM	dry ice; textile; clothing; paper product; greeting card; wrapping pa		
KM	fusion protein; ds.		
XX			
OS	Gaussia sp.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..1146	
FT		/+tag- a	
FT		/product- "Gaussia CBD-luciferase fusion protein"	
FT		/note- "CDS does not include start and stop codon"	
FT		/partial	
XX			
PN	WO200168824-A2.		
XX			
PD	20-SEP-2001.		
XX			
PF	15-MAR-2001; 2001WO-US08277.		
XX			
PR	15-MAR-2000; 2000US-189691P.		
XX			
PA	(PROL-) PROLUME LTD.		
PA	(BRYA/) BRYAN B J.		
XX			
PI	Bryan BJ, Szent-Gyorgyi C, Szczepaniak W;		
XX			
DR	WPI: 2002-010561/01.		
XX			
XX	P-PSDB: AAE13383.		
PT			
PR	Nucleic acids encoding Renilla reniformis green fluorescent proteins,		
XX	useful in diagnosis bioluminescence procedures -		
XX			
PS	Disclosure: Page 162-163; 17pp; English.		
XX			
XX			
CC	The patent discloses sea pansy (Renilla reniformis) green fluorescent		
CC	proteins (GFP) and their corresponding polynucleotides. The invention		
CC	also relates to sequences of the bioluminescence generating system		
CC	(e.g. luciferase). R. reniformis GFP are used in diagnostic methods		
CC	and in the production of novelty items such as toys (e.g. squirt gun,		
CC	pellet gun, toy "Halloween" eggs, toy cigarettes, board/card game		
CC	toys), finger paints, slimy play material, bubbles in bubble making		
CC	toys, fishing lures, dolls, sparklers, magic wand toys, balloons,		
CC	personal care item (e.g. cosmetic, bath powders, body creams, tooth		
CC	pastes, mouth wash, soaps, body paints, bubble bath), ornamental		
CC	transgenic plants, fountain, fairy dust, food (gelatin, icings,		
CC	frosting, beers, wine, champagne, milk, soft drinks, ice cubes, ice,		
CC	dry ice, beverage), textile (foot bag, clothing) and/or paper product		
CC	(greeting cards, wrapping paper). The present sequence is a DNA		


```

OS Piscirickettsia salmonis.
XX
PN CA2281913-A1.
XX
PD 17-MAR-2001.
XX
PF 17-SEP-1999; 99CA-2281913.
XX
PR 17-SEP-1999; 99CA-2281913.
XX
PA (KAYM/) KAY W W.
PA (BURI/) BURIAN J.
XX (KUZK/) KUZK M A.
XX
PI Kay WW, Burian J, Kuzk MA;
XX
DR WPI: 2001-316844/34.
XX
PT Method for protecting polkilothermic fish against salmonid rickettsial
PT septicemia and other rickettsial diseases comprises administering a
PT vaccine containing the OsPA protein of Piscirickettsia salmonis -
XX
XX
PS Example 3; Fig 4B; 35pp; English.
XX
XX
CC This invention relates to a method for the protection against infection
CC of a polkilothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed OsPA, or an immunogenic fragment of
CC OsPA in the form of a vaccine. The method is used for protecting animals,
CC particularly polkilothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis OsPA gene. The OsPA gene is used in the method of the
CC invention.
XX
SQ Sequence 118 BP; 21 A; 32 C; 28 G; 37 T; 0 other:

```

Query Match

```

Best Local Similarity 15.4%; Score 118; DB 22; Length 118;
Pred. No. 1e-24;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 493 GGTGAGAGCATGAGACGACGATATAATCAAACTGAACGAGTCTGGAAAAAGTGAA 552
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 118 GGTGAGAGCATGAGACGACGATATAATCAAACTGAACGAGTCTGGAAAAAGTGAA 59
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 553 GCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCGGTAACAGCTACTGTGTGAAC 610
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 58 GCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCGGTAACAGCTACTGTGTGAAC 1
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```

Search completed: October 27, 2002, 16:40:50
 Job time : 198.984 secs

THIS PAGE BLANK (USPTO)

RESULT 5	AR027448	AR027448	486 bp	DNA	linear	PAT 29-SEP-1999
LOCUS	AR027448	Sequence	1	from patent US 5856201.		
DEFINITION	AR027448	Accession				
ACCESSION	AR027448	Version				
VERSION	AR027448.1	GI	5938268			
KEYWORDS	Unknown.					
SOURCE	Unknown.					

ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 486)
AUTHORS	Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.
TITLE	Methods of detection using a cellulose binding fusion product
JOURNAL	Patent: US 5856201-A 1 05-JAN-1999;
FEATURES	Location/Qualifiers
source	1..486 /organism="unknown"
BASE COUNT	194 a 89 c 79 g 124 t
ORIGIN	
Query Match	35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity	98.9%; Pred. No. 3.2e-58;
Matches 272; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
QY	1 ATGTCAGTTGATTCCTACACCTCTAACCAATCGACAAACAACTCAATTACACCAATA 60
Db	16 ATGTCAGTTGATTTTACACACTCTAACCAATCGACAAACAACTCAATTACACCAATA 75
QY	61 ATCAAAATTTACTAACACATCTGCACAGTATTTAAATTTAAATGACGTAAAGTTGATAT 120
Db	76 ATCAAAATTTACTAACACATCTGCACAGTATTTAAATTTAAATGACGTAAAGTTGATAT 135
QY	121 TATTACACAGTATGATGTACACACAGGACAACTTCTGGTGTGACCAATGCTGGTCATTA 180
Db	136 TATTACACAGTATGATGTACACACAGGACAACTTCTGGTGTGACCAATGCTGGTCATTA 195
QY	181 TTAGCAAAATGATGTATGTGATTAACCTGACCAAGGACACAGCAAACTTGCTTAAAGAAC 240
Db	196 TTAGCAAAATGATGTATGTGATTAACCTGACCAAGGACACAGCAAACTTGCTTAAAGAAC 255
QY	241 GCAAGCCCAACATCAACCTATGATCATATCTGCA 275
Db	256 GCAAGCCCAACATCAACCTATGATCATATGTTGA 290
RESULT 6	
LOCUS	AR027449 486 bp DNA linear PAT 29-SEP-1999
DEFINITION	Sequence 3 from patent US 5856201.
ACCESSION	AR027449
VERSION	AR027449.1 GI:5938269
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 486)
TITLE	Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.
JOURNAL	Methods of detection using a cellulose binding fusion product
FEATURES	Patent: US 5856201-A 3 05-JAN-1999;
source	Location/Qualifiers
BASE COUNT	1..486 /organism="unknown"
ORIGIN	
Query Match	35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity	98.9%; Pred. No. 3.2e-58;
Matches 272; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
QY	1 ATGTCAGTTGATTCCTACACCTCTAACCAATCGACAAACAACTCAATTACACCAATA 60
Db	471 ATGTCAGTTGATTTTACACACTCTAACCAATCGACAAACAACTCAATTACACCAATA 412
QY	61 ATCAAAATTTACTAACACATCTGCACAGTATTTAAATTTAAATGACGTAAAGTTGATAT 120
Db	411 ATCAAAATTTACTAACACATCTGCACAGTATTTAAATTTAAATGACGTAAAGTTGATAT 352
QY	121 TATTACACAGTATGATGTACACAGGACAACTTCTGGTGTGACCAATGCTGGTCATTA 180

Db	351	TATTACAACTGATGTTGATACCAAGGACAAACTTTCTGCTGTTGACCACTGCTGTCATTA	292
QY	181	TTAGGAATAGCTATGTTGATTAACACTAGCAAAAGTGACGACAAACTTGGTTAAGAAACA	240
Db	291	TTAGGAATAGCTATGTTGATTAACACTAGCAAAAGTGACGACAAACTTGGTTAAGAAACA	232
QY	241	GCAAGCCCAACATCAACCTATGATACATATCTGGA	275
Db	231	GCAAGCCCAACATCAACCTATGATACATATCTTGA	197
RESULT 7			
LOCUS	AR058257	486 bp	DNA
DEFINITION	Sequence 1 from patent US 5837814.		linear
ACCESSION	AR058257		
VERSION	AR058257.1	GI:5983834	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 486)		
AUTHORS	Shoseyov,O., Shpiegl,I., Goldstein,M. and Dol,R.		
TITLE	Cellulose binding domain proteins		
JOURNAL	Patent: US 5837814-A 1 17-NOV-1998;		
FEATURES	Location/Qualifiers		
source	1..486		
BASE COUNT	194 a	89 c	79 g 124 t
ORIGIN	/organism="unknown"		
Query Match			
	Best Local Similarity	35.2%; Score 270.2; DB 6;	Length 486;
	Matches 272; Conservative	0; Mismatches 3;	Indels 0; Gaps 0;
QY	1	ATGCGAGTGGATCTTACTACAACTGACAAATCAGCAAAACAACTCAATTACACCAATA	60
Db	16	ATGTCAGTTGAATTTTACACTCTACAAATCAGCAAAACAACTCAATTACACCAATA	75
QY	61	ATCAAAATTTACTTAACACATCTGCAGAGTGATTTAAATTTAAATGACGTAAAGTTGATAT	120
Db	76	ATCAAAATTTACTTAACACATCTGCAGAGTGATTTAAATTTAAATGACGTAAAGTTGATAT	135
QY	121	TATTACACAGTGTGATGATACAAAGGACAAACTTCTGCTGTGACCACTGCTGTCATTA	180
Db	136	TATTACAAAGTGTGATGATACAAAGGACAAACTTCTGCTGTGACCACTGCTGTCATTA	195
QY	181	TTAGGAATAGCTATGTTGATTAACACTAGCAAAAGTGACGACAAACTTGGTTAAGAAACA	240
Db	196	TTAGGAATAGCTATGTTGATTAACACTAGCAAAAGTGACGACAAACTTGGTTAAGAAACA	255
QY	241	GCAAGCCCAACATCAACCTATGATACATATCTGGA	275
Db	256	GCAAGCCCAACATCAACCTATGATACATATCTTGA	290
RESULT 8			
LOCUS	AR058258	486 bp	DNA
DEFINITION	Sequence 3 from patent US 5837814.		linear
ACCESSION	AR058258		
VERSION	AR058258.1	GI:5983835	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 486)		
AUTHORS	Shoseyov,O., Shpiegl,I., Goldstein,M. and Dol,R.		
TITLE	Cellulose binding domain proteins		
JOURNAL	Patent: US 5837814-A 3 17-NOV-1998;		
FEATURES	Location/Qualifiers		
source	1..486		
	/organism="unknown"		

BASE COUNT 124 a 79 c 89 g 194 t
ORIGIN
Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 3.2e-58;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATTTCTCAACTCTTAACAATCAGCACAACCAAACTCAATTACACCAATA 60
DB 471 ATGTCAGTTGAATTTCTCAACTCTTAACAATCAGCACAACCAAACTCAATTACACCAATA 412
QY 61 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAGTAT 120
DB 411 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAGTAT 352
QY 121 TATTACACAAGTGATGTGTACACAAGACAACCTTCTGTGTGTGACCATGCTGTGTGCAATTA 180
DB 351 TATTACACAAGTGATGTGTACACAAGACAACCTTCTGTGTGTGACCATGCTGTGTGCAATTA 292
QY 181 TTAGGAATAGCTATGTTGATTAACACTAGCAAACTGACAGCAAACTTCGTTAAGAAACA 240
DB 291 TTAGGAATAGCTATGTTGATTAACACTAGCAAACTGACAGCAAACTTCGTTAAGAAACA 232
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
DB 231 GCAAGCCCAACATCAACCTATGATACATATCTGGA 197
RESULT 9
LOCUS 118514 486 bp DNA linear PAT 07-OCT-1996
DEFINITION Sequence 1 from patent US 5496934.
ACCESSION 118514.1 GI:1598869
VERSION 118514.1
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov,O., Shiplegl,I., Goldstein,M.A. and Doi,R.H.
TITLE Nucleic acids encoding a cellulose binding domain
JOURNAL Patent: US 5496934-A 1 05-MAR-1996;
FEATURES
source 1.486
location/Qualifiers
BASE COUNT 194 a 89 c 79 g 124 t
ORIGIN
Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 3.2e-58;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATTTCTCAACTCTTAACAATCAGCACAACCAAACTCAATTACACCAATA 60
DB 16 ATGTCAGTTGAATTTCTCAACTCTTAACAATCAGCACAACCAAACTCAATTACACCAATA 75
QY 61 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAGTAT 120
DB 76 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAGTAT 135
QY 121 TATTACACAAGTGATGTGTACACAAGACAACCTTCTGTGTGTGACCATGCTGTGTGCAATTA 180
DB 136 TATTACACAAGTGATGTGTACACAAGACAACCTTCTGTGTGTGACCATGCTGTGTGCAATTA 195
QY 181 TTAGGAATAGCTATGTTGATTAACACTAGCAAACTGACAGCAAACTTCGTTAAGAAACA 240
DB 196 TTAGGAATAGCTATGTTGATTAACACTAGCAAACTGACAGCAAACTTCGTTAAGAAACA 255
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
DB 256 GCAAGCCCAACATCAACCTATGATACATATCTGGA 290

RESULT 10
LOCUS 118515 486 bp DNA linear PAT 07-OCT-1996
DEFINITION Sequence 3 from patent US 5496934.
ACCESSION 118515
VERSION 118515.1 GI:1598870
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov,O., Shiplegl,I., Goldstein,M.A. and Doi,R.H.
TITLE Nucleic acids encoding a cellulose binding domain
JOURNAL Patent: US 5496934-A 3 05-MAR-1996;
FEATURES
source 1.486
location/Qualifiers
BASE COUNT 124 a 79 c 89 g 194 t
ORIGIN
Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 3.2e-58;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATTTCTCAACTCTTAACAATCAGCACAACCAAACTCAATTACACCAATA 60
DB 471 ATGTCAGTTGAATTTCTCAACTCTTAACAATCAGCACAACCAAACTCAATTACACCAATA 412
QY 61 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAGTAT 120
DB 411 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAGTAT 352
QY 121 TATTACACAAGTGATGTGTACACAAGACAACCTTCTGTGTGTGACCATGCTGTGTGCAATTA 180
DB 351 TATTACACAAGTGATGTGTACACAAGACAACCTTCTGTGTGTGACCATGCTGTGTGCAATTA 292
QY 181 TTAGGAATAGCTATGTTGATTAACACTAGCAAACTGACAGCAAACTTCGTTAAGAAACA 240
DB 291 TTAGGAATAGCTATGTTGATTAACACTAGCAAACTGACAGCAAACTTCGTTAAGAAACA 232
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
DB 231 GCAAGCCCAACATCAACCTATGATACATATCTGGA 197
RESULT 11
LOCUS 166614 486 bp DNA linear PAT 29-DEC-1997
DEFINITION Sequence 1 from patent US 5670623.
ACCESSION 166614
VERSION 166614.1 GI:2724592
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov,O., Shiplegl,I., Goldstein,M.A. and Doi,R.H.
TITLE Methods of use of cellulose binding domain proteins
JOURNAL Patent: US 5670623-A 1 23-SEP-1997;
FEATURES
source 1.486
location/Qualifiers
BASE COUNT 194 a 89 c 79 g 124 t
ORIGIN
Query Match 35.2%; Score 270.2; DB 6; Length 486;
Best Local Similarity 98.9%; Pred. No. 3.2e-58;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGTTGAATTTCTCAACTCTTAACAATCAGCACAACCAAACTCAATTACACCAATA 60
DB 16 ATGTCAGTTGAATTTCTCAACTCTTAACAATCAGCACAACCAAACTCAATTACACCAATA 75

Oy	61	ATCAAAATTACTACACACTGACAGTGATTTAAATTTAAATGACGTAAAGTTGAGAT	120
Db	76	ATCAAAATTACTAACACATCTGCACAGTATTTAAATTTAAAGACGTAAAGCTTAAAGATAT	135
Oy	121	TATTACACAAGTGATGTGACACAAAGACAATTTCTGTGTGACACAGTCGTGCATTA	180
Db	136	TATTACACAAGTGATGTGACACAAAGACAATTTCTGTGTGACACAGTCGTGCATTA	195
Oy	181	TTTAGCAAAATAGCTATGTTGATTAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAACA	240
Db	196	TTAGCAAAATAGCTATGTTGATTAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAACA	255
Oy	241	GCAAGCCCAACATCAACCGTATGATCATATATCGGA	275
Db	256	GCAAGCCCAACATCAACCGTATGATCATATATGTTGA	290

RESULT	12			
LOCUS	166615/c			
DEFINITION	Sequence 3 from patent US 5670623.	486 bp	DNA	linear
ACCESSION	166615			PAT 29-DEC-1997
VERSION	166615.1	GI:2724593		
KEYWORDS				
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 486)			
TITLE	Shoseyov, O., Shpiegl, I., Goldstein, M. A. and Doi, R. H.			
JOURNAL	Methods of use of cellulose binding domain proteins			
FEATURES	Patent: US 5670623-A 3 23-SEP-1997;			
source	Location/Qualifiers			
	1..486			

BASE COUNT	124 a	79 c	89 g	194 t
ORIGIN	/organism="unknown"			

Query Match	35.2%	Score 270.2	DB 6	Length 486
Best Local Similarly	98.9%	Pred. No. 3.2e-58		
Matches 272; Conservative	0	Mismatches 3	Indels 0	Gaps 0

OY	1	ATTCAGTGTGAATTCCTCAACCTCTAACCAATCAGCACAAACAAACCTAATTTACACCAAT	60
Db	471	ATGTAGTGTGAATTTTACAACTCTACCAATCAGCACAAACAAACCTAATTTACACCAAT	412
OY	61	ATCAAAATTTACTAACACATCTGACAGTGAATTTAAATTTAAATGACGTAAAGTTAGTAT	120
Db	411	ATCAAAATTTACTAACACATCTGACAGTGAATTTAAATTTAAATGACGTAAAGTTAGTAT	352
OY	121	TATTTACACAAGTGAAGGTGATACCAAGGACAACTTTGCGGTGACACTGGTGGCATTA	180
Db	351	TATTTACACAAGTGAAGGTGATACCAAGGACAACTTTGCGGTGACACTGGTGGCATTA	292
OY	181	TTAGGAATATAGCTATGTTGATATACACTAGCAAACTGACAGCAAACTTGTTAAAGAAACA	240
Db	291	TTAGGAATATAGCTATGTTGATATACACTAGCAAACTGACAGCAAACTTGTTAAAGAAACA	232
OY	241	GCAAGCCCAACATCAACCTATGATATCATTTGGA	275
Db	231	GCAAGCCCAACATCAACCTATGATATCATTTGGA	197

RESULT 13			
188789			
LOCUS	188789	486 bp	DNA
DEFINITION	Sequence 1 from patent US 5719044.		linear
ACCESSION	188789		PAT 10-AUG-1998
VERSION	188789.1	GI:3408729	
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
	1 (bases 1 to 486)		

```

AUTHORS      Shoseyov,O., Shpiegel,I., Goldstein,M.M. and Dol,R.H
TITLE        Cellulose binding domain fusion proteins
JOURNAL      Patent: US 5719044-A 17-FEB-1998;
FEATURES     Location/Qualifiers
              source
                1..486
                /organism="unknown"
BASE COUNT   194 a      89 c      79 g      124 t
ORIGIN

```

Query Match	35.2%	Score 270.2	DB 6	Length 486
Best Local Similarity	98.9%	Pred. No. 3.2e-58		
Matches 272; Conservative	0	Mismatches 3	Indels 0	Gaps 0

QY	1	ATTCAGTTGAAATCTACAACTGTAAACAAATCACCACAAACAAACCTAATTACCCAAAT	60
Db	16	ATGTAGTTGAAATTTTACAACTGTACAAATCAGACACAAACAACTAATTACCCAAAT	75
QY	61	ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT	120
Db	76	ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT	135
QY	121	TATTACACAAGTGAAGTATACAAAGGCAAACTTGTGGTGTGACCATGGTGGTCATAT	180
Db	136	TATTACACAAGTGAAGTATACAAAGGCAAACTTGTGGTGTGACCATGGTGGTCATAT	195
QY	181	TTAGGAATTAGCTATGTGATATACACTAGCAAGTGAAGCAAACTTCGTTAAAGAAAC	240
Db	196	TTAGGAATTAGCTATGTGATATACACTAGCAAGTGAAGCAAACTTCGTTAAAGAAAC	255
QY	241	GCAGCCCAACATCACTATATGATATCATATGTGGA	275
Db	256	GCAGCCCAACATCACTATATGATATCATATGTGGA	290

RESULT 14
I88790/c

LOCUS	188790	486 bp	DNA	Linear	PAT 10-AUG-1998
DEFINITION	Sequence 3 from patent US 5719044.				
ACCESSION	188790				
VERSION	188790.1	GI:3408730			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 486)				
TITLE	Shoseyov, O., Shpiegl, I., Goldstein, M. A. and Doi, R. H.				
JOURNAL	Cellulose binding domain fusion proteins				
FEATURES	Patent: US 5719044-A 3 17-FEB-1998;				
source	Location/Qualifiers				
	1..486				

	/organism="unknown"			
BASE COUNT	124 a	79 c	89 g	194 t
ORIGIN				

Query Match	35.2%	Score 270.2	DB 6	Length 486
Best Local Similarity	98.9%	Pred. No. 3.2e-58		
Matches 27; Conservative	0	Mismatches 3	Indels 0	Gaps 0

Qy	1	ATGTCAGTTGATTTGTACAACTCTAACCAATGACACAAACAACTCAATTACCCGATA	60
Db	471	ATGTCAGTTGATTTGTACAACTCTAACCAATGACACAAACAACTCAATTACCCGATA	412
Qy	61	ATCAAAATTTACTAACACATCTGCACAGTATTTAAATTTAAATGACGTAAAGTTAGATAT	120
Db	411	ATCAAAATTTACTAACACATCTGCACAGTATTTAAATTTAAATGACGTAAAGTTAGATAT	352
Qy	121	TATTCACACAGTGTGATACACAAAGGACAACTTTCTGGTGTACCAATGCTGGTGCATTA	180
Db	351	TATTCACACAGTGTGATACACAAAGGACAACTTTCTGGTGTACCAATGCTGGTGCATTA	292
Qy	181	TTAGGAAATAGCTATGTTGATATACACTGTGCAAAATGACACGAACCTCGTTAAAGAAACA	240
Db	291	TTAGGAAATAGCTATGTTGATATACACTGTGCAAAATGACACGAACCTCGTTAAAGAAACA	322

QY 241 GCAAGCCCAACATCACTATGATACATATCTGGA 275
|||||
Db 231 GCAAGCCCAACATCACTATGATACATATGTTGA 197

RESULT 15

AR096204 499 bp DNA linear PAT 08-SEP-2000

LOCUS AR096204 Sequence 5 from patent US 6005092.

ACCESSION AR096204

VERSION AR096204.1 GI:10024795

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 272: Conservative

0: Mismatches

3: Indels

0: Gaps

0:

QY 1 ATGTGAGTTGAATTTACAACTCTACAAATCAGCAAAACAACATCAATTACACCATTA 60

Db 21 ATGTGAGTTGAATTTACAACTCTACAAATCAGCAAAACAACATCAATTACACCATTA 80

QY 61 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATTAATGACGTAAAGTTAGATAT 120

Db 81 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATTAATGACGTAAAGTTAGATAT 140

QY 121 TATTACACAAGTGATGTGACACAAGCAAACTTCTGCTGATGACATGCTGTCATTA 180

Db 141 TATTACACAAGTGATGTGACACAAGCAAACTTCTGCTGATGACATGCTGTCATTA 200

QY 181 TTAGAAATAGCTATGTGATACACTAGCAAGTAGCAAACTTCTGTTAAGAAACA 240

Db 201 TTAGAAATAGCTATGTGATACACTAGCAAGTAGCAAACTTCTGTTAAGAAACA 260

QY 241 GCAAGCCCAACATCACTATGATACATATCTGGA 275

Db 261 GCAAGCCCAACATCACTATGATACATATGTTGA 295

Search completed: October 27, 2002, 17:46:49
Job time : 1495.36 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 20:15:49 : Search time 1538.65 Seconds
(without alignments)
6736.842 Million cell updates/sec

Title: US-09-677-374-5

Perfect score: 768

Sequence: 1 atcgtcagttgaattctacaa.....agggtgattgacacgcgaataaa 768

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST :
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlun:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	2.9	709	9	AV760726
2	21	2.7	391	12	AQ252889
3	20	2.6	516	12	AQ518147
4	20	2.6	717	10	BE532389
5	20	2.6	823	10	BI186311
6	19	2.5	184	10	BE247681
7	19	2.5	190	12	FR0011976
8	19	2.5	191	9	AM661978
9	19	2.5	210	9	BF803712
10	19	2.5	237	9	AA909847
11	19	2.5	265	10	Z44341
12	19	2.5	283	9	A1817750
13	19	2.5	322	9	A1182462
14	19	2.5	335	10	DS1829
15	19	2.5	336	10	BE246888
16	19	2.5	348	10	T51729
17	19	2.5	363	9	AA825376

18	19	2.5	384	10	BF109466	BF109466 7152d12.x
19	19	2.5	456	9	A1559589	A1559589 t650h10.x
20	19	2.5	459	9	AV763598	AV763598 AV763598
21	19	2.5	475	9	AM068733	AM068733 cn24c11.x
22	19	2.5	502	10	BM309775	BM309775 sak6e03.
23	19	2.5	508	9	A1478023	A1478023 fb50608.x
24	19	2.5	510	9	A1678459	A1678459 t82f07.x
25	19	2.5	527	10	BM381932	BM381932 MEST542-A
26	19	2.5	528	9	AM658421	AM658421 94440 MAR
27	19	2.5	536	9	A1637152	A1637152 603001D08
28	19	2.5	630	10	BM335231	BM335231 MEST147-F
29	19	2.5	645	10	BC924401	BC924401 HNC26-1-H
30	19	2.5	648	12	BH381604	BH381604 AG-ND-132
31	19	2.5	652	9	A1267409	A1267409 ag64e09.x
32	19	2.5	670	10	BI151293	BI151293 602917270
33	19	2.5	678	9	AM167360	AM167360 x974h05.x
34	19	2.5	697	10	BM350333	BM350333 MEST264-C
35	19	2.5	719	10	BI186832	BI186832 603392552
36	19	2.5	744	10	BI764273	BI764273 603045965
37	19	2.5	816	12	BH524872	BH524872 BCGFY55FE
38	19	2.5	846	10	BI667298	BI667298 603292555
39	19	2.5	892	9	AU135371	AU135371 AU135371
40	19	2.5	919	12	CNS04Q0J	AL303076 Tetradon
41	19	2.5	2405	10	AV214643	AV214643 602673281
42	18	2.3	342	12	CNS005SW	AV214643 AV214643
43	18	2.3	372	12	BH395820	AL088814 Arabidops
44	18	2.3	382	9	AM588379	BH395820 AG-ND-119
45	18	2.3	382	9	AM588379	AM588379 pa12d10.y

ALIGNMENTS

RESULT 1
LOCUS AV760726 709 bp mRNA linear EST 19-OCT-2000
DEFINITION AV760726 MDS Homo sapiens cDNA clone MDSAJE03 5', mRNA sequence.
ACCESSION AV760726
VERSION AV760726.1 GI:10918574
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 709)
AUTHORS Gu,J., Zhao,M., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,H., Xiao,H., Gu,Y., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu,M., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,Y., Gao,G., Zhang,O., Chen,S., Han,Z. and Chen,Z.
Homo sapiens cDNA MDS clones
Unpublished (2000)
JOURNAL
TITLE
COMMENT
CONTACT: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
LOCATION/Qualifiers
1. 709
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MDSAJE03"
/clone_lib="MDS"
/tissue_type="Bone marrow"
/cell_type="CD34+ hematopoietic stem/progenitor cell"
/lab_host="BM25.8"
/note="Vector: pTrioPE2; Site_1: sf11a; Site_2: sf11b"

BASE COUNT 172 a 159 c 188 g 190 t
ORIGIN
Query Match 2.9%; Score 22; DB 9; Length 709;

Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 147 ACAACTTCTGTGTCACCAT 168
|||||
Db 497 ACAACTTCTGTGTCACCAT 518

RESULT 2
AQ252889/c 391 bp DNA linear GSS 10-OCT-1998
LOCUS HS_3029_A2.D03_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3029 Col=6 Row=G, DNA sequence.
ACCESSION AQ252889
VERSION AQ252889.1 GI:3723803
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 391)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3029 row: G column: 6
Class: BAC ends
High quality sequence stop: 391.
Location/Qualifiers
1..391
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3029 Col=6 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 77 a 73 c 85 g 156 t

ORIGIN

Query Match 2.7%; Score 21; DB 12; Length 391;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 AATCAGCACAACAACACTCA 49
|||||
Db 153 AATCAGCACAACAACACTCA 133

RESULT 3
AO518147 516 bp DNA linear GSS 05-MAY-1999
LOCUS HS_5090_A1.F08_T7A RPTC-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=666 Col=15 Row=K, DNA sequence.
ACCESSION AO518147
VERSION AO518147.1 GI:4743329
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 516)

AUTHORS Mahairas G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPTC-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (Info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 666 row: K column: 15
Seq primer: 17
Class: BAC ends
High quality sequence stop: 516.
Location/Qualifiers
1..516
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=666 Col=15 Row=K"
/clone_lib="RPTC-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

BASE COUNT 122 a 156 c 129 g 104 t 5 others

ORIGIN

Query Match 2.6%; Score 20; DB 12; Length 516;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 296 GCCTCGACGGCAGCTCTCTG 315
|||||
Db 191 GCCTCGACGGCAGCTCTCTG 210

RESULT 4
BE532389 717 bp mRNA linear EST 09-AUG-2000
LOCUS 601233730F1 NCL_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3597462 5'
DEFINITION mRNA sequence.
ACCESSION BE532389
VERSION BE532389.1 GI:9761034
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 717)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
http://image.llnl.gov

Plate: LLM8776 row: 1 column: 07
High quality sequence stop: 510.
Location/Qualifiers
1. 717

FEATURES
source
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3597462"
/clone_1lb="NCI CGAP Mam6"
/sex="female, virgin"
/tissue_type="Infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: Salt; Site:2: NotI. Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT 228 a 152 c 150 g 187 t
ORIGIN

Query Match 2.68; Score 20; DB 10; Length 717;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 26 ACAATCAGCACAACAAC 45
|||||
Db 462 ACAATCAGCACAACAAC 481

RESULT 5
B1186311/c
LOCUS B1186311 823 bp mRNA linear EST 10-JUL-2001
DEFINITION UNL-P-FN-cf-b-07-0-UNL.52 UNL-P-FN Sus scrofa cDNA clone
UNL-P-FN-cf-b-07-0-UNL 3', mRNA sequence.
ACCESSION B1186311
VERSION B1186311.1 GI:14660720
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE
AUTHORS Caelano,A.R., Johnson,R.K. and Pomp,D.
TITLE Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles
JOURNAL unpublished (2001)
COMMENT Contact: Pomp, D
Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomp@unl.edu
Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message.
Seq primer: M13 -29

FEATURES
source
Location/Qualifiers
1. 823
/organism="Sus scrofa"
/strain="University of Nebraska, Lincoln Swine Selection
Lines"
/db_xref="taxon:9823"
/clone="UNL-P-FN-cl-b-07-0-UNL"
/clone_1lb="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-pac (Pharmacia) with a modified
polylinker; Site:1: Not I; Site:2: Eco RI; The UNL-P-FN
library is a normalized library representing porcine
ovarian follicles, ranging between 2.0 to 10.0 mm in
diameter, collected during 7 days of the follicular phase
of the pig estrous cycle. This library was derived from

BASE COUNT 151 a 272 c 231 g 167 t 2 others
ORIGIN

Query Match 2.68; Score 20; DB 10; Length 823;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 293 GTTGCCTGACGACGACTCT 312
|||||
Db 730 GTTGCCTGACGACGACTCT 711

RESULT 6
BE247681/c
LOCUS BE247681 184 bp mRNA linear EST 03-OCT-2001
DEFINITION TCBAP2E3536 Pediatric pre-B cell acute lymphoblastic leukemia
Baylor-HGSC project-TCBA Homo sapiens cDNA clone TCBAP3536, mRNA
sequence.
ACCESSION BE247681
VERSION BE247681.1 GI:9099528
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Wei,X., Tsang,X.T.M., Mei,G., Ku,J.M., Ali-Osman Jr.,F.R., Muzny,D.,
Bouck,J., Gibbs,R.A. and Margolin,J.F.
TITLE Pediatric Leukemia cDNA Sequencing Project
JOURNAL unpublished (2000)
COMMENT Contact: Dr. Judith F. Margolin
Texas Children's Cancer Center and Human Genome Sequencing Center
at Baylor College of Medicine
1102 Bates, MC3-3320 Houston, TX 77030, USA
Tel: 832-824-4536
Fax: 832-825-4038
Email: clones@ccc.org

Citation: Carninci,P. and Hayashizaki,Y. High efficiency
full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Seq primer: M13 primer.
Location/Qualifiers
1. 184
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TCBAP3536"
/clone_1lb="Pediatric pre-B cell acute lymphoblastic
leukemia Baylor-HGSC project-TCBA"
/sex="male"
/tissue_type="leukopheresis"
/cell_type="pre-B cell"
/dev_stage="pediatric 2 years"
/lab_host="DH10B"
/note="Vector: lambda PSB; Site:1: BamHI; Site:2: EcoRI;
First strand cDNA was primed with an anchored
XhoI-oligo(dT) primer [5'GGAGAGCTCGAGCGCGCGAGAG(T)VN
3'; V=A,C,G, N=A,C,G,T] and then dg tailed. Second strand
was primed with a BamHI-dc primer
[5'AGAGAGCTCGAGCTCGCGCGCGCAATTAATAT(C) 3'].
Double-stranded cDNA was then digested with BamHI and XhoI
and directionally cloned into the BamHI and SalI sites of
lambda psb vector. Library went through one round of
normalization. Library was constructed by Wei Yu at RIKEN
of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,
Itoh M, Nagaoka S, Sasaki N, Okazaki Y, Muramatsu M,
Schneider C, Hayashizaki Y, High efficiency selection of
full-length cDNA by improved biotinylated cap trapper.,
DNA Res 4: 1, 61-66, Feb 28, 1997)."

FEATURES
source

BASE COUNT 43 a 27 c 30 g 79 t 5 others
 ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 184;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 ACAGTGATTTAAATTAA 101
 ||||||||||||||||
 Db 31 ACAGTGATTTAAATTAA 13

RESULT 7
 FR0011976
 LOCUS 190 bp DNA linear GSS 18-SEP-1997
 DEFINITION F.rubripes GSS sequence, clone 072L05AE5, genomic survey sequence.
 ACCESSION AF003232
 VERSION AF003232.1 GI:2448802
 KEYWORDS GSS; genome survey sequence.
 SOURCE Takifugu rubripes.
 ORGANISM Takifugu rubripes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 Tetraodontidae; Takifugu.
 1 (bases 1 to 190)
 Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y.,
 Williams,G. and Brenner,S.
 Direct Submission
 Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource
 Centre Hinxton, Cambridge, CB10 1SB. Email: biolhelp@hmp.mrc.ac.uk
 Vector: m13mp18
 V-type: phage
 PRIMER: M13
 DESC: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

FEATURES
 source Location/Qualifiers
 1..190
 /organism="Takifugu rubripes"
 /db_xref="taxon:31033"
 /clone_id="cosmid 072L05"
 /clone="072L05AE5"

BASE COUNT 56 a 38 c 40 g 56 t
 ORIGIN

Query Match 2.5%; Score 19; DB 12; Length 190;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 CAGGCAGCTCTGATCA 319
 ||||||||||||||||
 Db 119 CAGGCAGCTCTGATCA 137

RESULT 8
 AW661978
 LOCUS 191 bp mRNA linear EST 06-APR-2000
 DEFINITION h117h11.x1 NCI-CGAP-GUI Homo sapiens CDNA clone IMAGE:2972613 3',
 similar to contains element TARI repetitive element ;, mRNA
 sequence.
 ACCESSION AW661978
 VERSION AW661978.1 GI:7454514
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 191)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)

COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life
 Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University
 Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 image.llnl.gov/image/html/lresources.shtml
 Seq primer: -400p from Glbco
 High quality sequence stop: 157.

FEATURES
 source Location/Qualifiers
 1..191
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2972613"
 /clone_id="NCI-CGAP-GUI"
 /tissue_type="2 pooled high-grade transitional cell
 tumors"
 /lab_host="DH10B"
 /note="Organ: genitourinary tract; Vector: pCMV-SPORT6;
 Site.1: SalI; Site.2: NotI; Cloned unidirectionally.
 Primer: oliigo dt. Library constructed by Life
 Technologies."

BASE COUNT 70 a 32 c 35 g 54 t
 ORIGIN

Query Match 2.5%; Score 19; DB 9; Length 191;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 ACAGTGATTTAAATTAA 101
 ||||||||||||||||
 Db 32 ACAGTGATTTAAATTAA 50

RESULT 9
 BF803712
 LOCUS 210 bp mRNA linear EST 12-JAN-2001
 DEFINITION CM0-C10136-021100-670-d11 C10136 Homo sapiens CDNA, mRNA sequence.
 ACCESSION BF803712
 VERSION BF803712.1 GI:12132701
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 210)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R.,
 Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM0&cl=CM0-C10136-
 021100-670-d11&ts=2000-11-02&td=1)
 Seq primer: puc 18 forward
 High quality sequence start: 6

FEATURES High quality sequence stop: 210.
Location/Qualifiers
1. 210
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1ib="C10136"
/dev_stage="Adult"
/note="Organ: colon.ins; Vector: puc18; Site.1: SmaI;
Site.2: SmaI. A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 95 a 35 c 24 g 55 t 1 others

ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 210;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 83 ACAGTGATTTAAATTAA 101
|||||
Db 174 ACAGTGATTTAAATTAA 192

RESULT 10 237 bp mRNA linear EST 26-AUG-1998
AA909847
LOCUS 0101910.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1522242 3',
DEFINITION mRNA sequence.
ACCESSION AA909847
VERSION AA909847.1 GI:3047549
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 237)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs1@mail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LINL at:
www.bio.lnll.gov/bdrip/image/image.html
Insert Length: 528 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 210.
Location/Qualifiers
1. 237
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1522242"
/clone_1ib="NCI_CGAP_Lu5"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker. 1st strand cDNA was prepared from
neuroendocrine lung carcinoid, and was then primed with a
Not I - oligo(dT) primer. Double-stranded cDNA was ligated
to Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library is normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 91 a 41 c 42 g 63 t

ORIGIN

Query Match 2.5%; Score 19; DB 9; Length 237;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 83 ACAGTGATTTAAATTAA 101
|||||
Db 31 ACAGTGATTTAAATTAA 49

RESULT 11 244341
244341/c
LOCUS HSC1YD071 normalized infant brain cDNA Homo sapiens cDNA clone
DEFINITION c-1y407, mRNA sequence.
ACCESSION 244341
VERSION 244341.1 GI:573469
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 265)
Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes
M.D., Duprat, S., Houlgate, R., Juneau, M.N., Lamy, B., Lorenzo, F.,
Mitchell, H., Marage-Samson, R., Pietu, G., Pouliot, Y.,
Sebastiani, Kabakthis, C. and Tessier, A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
95277534
Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress library_id: C; Genexpress_sequence_id: y1c-1y407
Seq primer: (-21)M13-universal.
Location/Qualifiers
1. 265
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-1y407"
/clone_1ib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site.1: HindIII;
Site.2: NotI; sex=Female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue_type-total brain
; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B. Soares, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"

BASE COUNT 76 a 43 c 48 g 89 t 9 others

ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 265;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 83 ACAGTGATTTAAATTAA 101
|||||
Db 209 ACAGTGATTTAAATTAA 191

RESULT 12
A1817750

LOCUS A1817750 283 bp mRNA linear EST 21-DEC-1999
 DEFINITION wk25909.x1 NCI-CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2413408 3',
 mRNA sequence.
 ACCESSION A1817750
 VERSION A1817750.1 GI:5436829
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 283)
 NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute / National Institute of Neurological
 Disorders and Stroke, Brain Tumor Genome Anatomy Project
 (CGAP/BRGAP), Tumor Gene Index
 Unpublished (1998)
 Contact: Robert Strausberg, Ph.D.
 Email: cgap@r-remail.nih.gov
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
 Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbrrp/image/image.html
 Insert length: 346 Std Error: 0.00
 Seq primer: -40up from GIBCO.
 Location/Qualifiers
 1..283
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2413408"
 /clone_1lb="NCI CGAP_Brn25"
 /tissue_type="anaplastic oligodendroglioma"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer (5'
 TGTTCACATCTGACGAGCGGCGCGCATGCTTTTGTGTGTGTGTGTGTGT
 T 3'); double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pT73 vector.
 Library is normalized, and was constructed by Bento
 Soares and M.Fatima Bonaldo."

BASE COUNT 96 a 62 c 46 g 79 t
 ORIGIN
 Query Match 2.5%; Score 19; DB 9; Length 283;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 ACAGTATTAAATTAA 101
 ||||||||||||||||
 Db 36 ACAGTATTAAATTAA 54

RESULT 13
 A1182462 322 bp mRNA linear EST 08-OCT-1998
 LOCUS uc26h10.f1 Soares_mammary_gland_NbMNG Mus musculus cDNA clone
 DEFINITION IMAGE:1399171 5', mRNA sequence.
 ACCESSION A1182462
 VERSION A1182462.1 GI:3733100
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 322)
 Mammaia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Mus.
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

TITLE Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 The Washu-HMI Mouse EST Project
 Unpublished (1996)
 CONTACT: Mairra M/Mouse EST Project
 Washu-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:910887
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 303.
 Location/Qualifiers
 1..322
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:1399171"
 /clone_1lb="Soares_mammary_gland_NbMNG"
 /sex="male"
 /tissue_type="mammary gland"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)
 primer (5'
 TGTTCACATCTGACGAGCGGCGCGCATGCTTTTGTGTGTGTGTGTGT
 T 3'); double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pT73 vector.
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
 constructed and normalized by Bento Soares and M.Fatima
 Bonaldo."

BASE COUNT 47 a 100 c 95 g 79 t 1 others
 ORIGIN
 Query Match 2.5%; Score 19; DB 9; Length 322;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 TGGTACACAGCAAACT 153
 ||||||||||||||||
 Db 185 TGGTACACAGCAAACT 167

RESULT 14
 D51829 335 bp mRNA linear EST 14-SEP-1995
 LOCUS HUM053G02B Clontech human fetal brain polyA+ mRNA (#535) Homo
 DEFINITION sapiens cDNA clone GEN-053G02 5', mRNA sequence.
 ACCESSION D51829
 VERSION D51829.1 GI:952065
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 335)
 Fujikawa, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,
 Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Takachi,
 A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H.,
 Shin, S. and Nakamura, Y.
 Fujikawa et al. (1995)
 Unpublished (1995)
 Contact: Tsutomu Fujikawa
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co., Ltd

463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035

FEATURES

source

Location/Qualifiers

1. .335

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="GEN-053602"

/clone_lib="Clontech human fetal brain polyA+ mRNA (#6535)

BASE COUNT 93 a 44 c 58 g 126 t 14 others

ORIGIN

Query Match 2.5%: Score 19; DB 10; Length 335;

Best Local Similarity 100.0%; Pred. No. 55;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 83 ACAGTGATTTAAATTAA 101

Db 95 ACAGTGATTTAAATTAA 77

RESULT 15

BE246888 336 bp mRNA linear EST 03-OCT-2001

LOCUS TCBAP15756 Pediatric pre-B cell acute lymphoblastic leukemia

DEFINITION Baylor-HGSC project-TCBA Homo sapiens cDNA clone TCBAP5756; mRNA

sequence.

ACCESSION BE246888

VERSION BE246888.1 GI:9098639

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (Bases 1 to 336)

Wei Y., Tsang Y.T.M., Mei G., Ku J.M., Ali-Osman Jr., F.R., Muzny, D.

, Bouck J., Gibbs R.A. and Margolin, J.F.

Pediatric Leukemia cDNA Sequencing Project

Unpublished (2000)

Contact: Dr. Judith F. Margolin

Texas Children's Cancer Center and Human Genome Sequencing Center

at Baylor College of Medicine

1102 Bates, MC3-3320 Houston, TX 77030, USA

Tel: 832-824-4536

Fax: 832-825-4038

Email: clones@xccc.org

Citation: Carninci, P. and Hayashizaki, Y. High efficiency

full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

Seq primer: M13 primer.

Location/Qualifiers

1. .336

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="TCBAP5756"

/clone_lib="Pediatric pre-B cell acute lymphoblastic

leukemia Baylor-HGSC project-TCBA"

/sex="male"

/tissue_type="leukopheresis"

/cell_type="pre-B cell"

/dev_stage="pediatric 2 years"

/lab_host="DH108"

/note="Vector: lambda PSB; Site_1: BamHI; Site_2: EcoRI;

First strand cDNA was then digested with BamHI and XhoI

and directionally cloned into the BamHI and SalI sites of

lambda PSB vector. Library went through one round of

normalization. Library was constructed by Wei Yu at RIKEN

of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,

Ichih M, Nagaoka S, Sasaki, Okazaki Y, Muramatsu M,
Schneider C, Hayashizaki Y, High efficiency selection of
full-length cDNA by improved biotinylated cap trapper.,
DNA Res 4: 1, 61-6, Feb 28, 1997)

BASE COUNT

99 a 43 c 56 g 138 t

ORIGIN

Query Match 2.5%: Score 19; DB 10; Length 336;

Best Local Similarity 100.0%; Pred. No. 55;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 83 ACAGTGATTTAAATTAA 101

Db 31 ACAGTGATTTAAATTAA 13

Search completed: October 27, 2002, 22:36:47
Job time: 1544.65 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 20:19:49 : Search time 57.0363 Seconds
(without alignments)
3307.484 Million cell updates/sec

Title: US-09-677-374-5

Perfect score: 768
Sequence: 1 atgtcagttgattcttctacaa.....aggtgattagcaccgaaaaa 768

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size: 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Issued Patents.NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCrUS_COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	255	33.2	486	1 US-08-048-164A-1	Sequence 1, Appl1
2	255	33.2	486	1 US-08-048-164A-3	Sequence 3, Appl1
3	255	33.2	486	1 US-08-460-462-1	Sequence 1, Appl1
4	255	33.2	486	1 US-08-460-462-3	Sequence 3, Appl1
5	255	33.2	486	1 US-08-460-457-1	Sequence 1, Appl1
6	255	33.2	486	1 US-08-460-457-3	Sequence 3, Appl1
7	255	33.2	486	1 US-08-460-458-1	Sequence 1, Appl1
8	255	33.2	486	1 US-08-460-458-3	Sequence 3, Appl1
9	255	33.2	486	2 US-08-460-455-1	Sequence 1, Appl1
10	255	33.2	486	2 US-08-460-455-3	Sequence 3, Appl1
11	255	33.2	486	2 US-08-330-394A-1	Sequence 1, Appl1
12	255	33.2	486	2 US-08-330-394A-3	Sequence 3, Appl1
13	255	33.2	486	2 US-09-006-636-5	Sequence 5, Appl1
14	255	33.2	486	2 US-09-006-632-5	Sequence 5, Appl1
15	255	33.2	486	2 US-09-277-716-21	Sequence 21, Appl1
16	255	33.2	486	2 US-08-330-394A-23	Sequence 23, Appl1
17	19	2.6	26	2 US-08-198-956-9	Sequence 9, Appl1
18	19	2.5	1482	4 US-09-198-955A-11	Sequence 11, Appl1
19	18	2.3	3808	2 US-08-916-917-3	Sequence 3, Appl1
20	18	2.3	3808	2 US-08-972-631-3	Sequence 3, Appl1
21	18	2.3	3808	2 US-08-972-629-3	Sequence 3, Appl1
22	18	2.3	3808	2 US-08-972-630-3	Sequence 3, Appl1
23	18	2.3	3808	2 US-08-672-211-3	Sequence 3, Appl1
24	18	2.3	3808	2 US-09-225-170-3	Sequence 3, Appl1
25	18	2.3	3808	1 US-08-074-879-9	Sequence 9, Appl1
26	18	2.3	7833	1 US-08-468-057A-9	Sequence 9, Appl1
27	18	2.3	4411529	4 US-09-103-840A-1	Sequence 1, Appl1

c 28	17	2.2	754	4 US-09-541-941B-26	Sequence 26, Appl1
c 29	17	2.2	1384	1 US-08-553-888A-1	Sequence 1, Appl1
c 30	17	2.2	1466	1 US-08-553-888A-2	Sequence 2, Appl1
c 31	17	2.2	4982	3 US-08-599-103B-1	Sequence 1, Appl1
c 32	17	2.2	4982	4 US-09-229-059-1	Sequence 1, Appl1
c 33	16	2.1	88	4 US-09-042-071-49	Sequence 49, Appl1
c 34	16	2.1	707	2 US-08-465-380-37	Sequence 37, Appl1
c 35	16	2.1	707	2 US-08-486-397-37	Sequence 37, Appl1
c 36	16	2.1	707	2 US-08-486-399-37	Sequence 37, Appl1
c 37	16	2.1	707	2 US-08-461-965-37	Sequence 37, Appl1
c 38	16	2.1	707	2 US-08-634-641-37	Sequence 37, Appl1
c 39	16	2.1	707	3 US-09-249-471-37	Sequence 37, Appl1
c 40	16	2.1	707	3 US-09-249-472-37	Sequence 37, Appl1
c 41	16	2.1	707	3 US-09-249-451-37	Sequence 37, Appl1
c 42	16	2.1	707	3 US-08-809-455-37	Sequence 37, Appl1
c 43	16	2.1	707	3 US-09-249-461-37	Sequence 37, Appl1
c 44	16	2.1	707	3 US-09-249-448-37	Sequence 37, Appl1
c 45	16	2.1	787	4 US-08-858-207A-210	Sequence 210, Appl1

ALIGNMENTS

RESULT 1
US-08-048-164A-1
Sequence 1, Application US/08048164A
Patent No. 5496934
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shlegel, Itai
APPLICANT: Goldstein, Marc A.
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/048,164A
FILING DATE: 14-Apr-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REFERENCE/DOCKET NUMBER: 18,872
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..486
US-08-048-164A-1
Query Match 33.2%; Score 255; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 5.5e-124;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	16	TACAACTCTAAACCAATGACGCCAAACAAACGCAATTACACCAATATATCAAAATTACTAAC	75
Db	31	TACAACCTCTAAACCAATGACGCCAAACAAACGCAATTACACCAATATATCAAAATTACTAAC	90
OY	76	ACAACTGCAGCTGATTTAAATTTAAATGACGTAAAGCTTGAATTTATTCACAAAGTAT	135
Db	91	ACAACTGCAGCTGATTTAAATTTAAATGACGTAAAGCTTGAATTTATTCACAAAGTAT	150
OY	136	GGTACACAGACCAACACTTTCTGCTGTGACCATCTGCTGCATTATTAGGAATAGCTAT	195
Db	151	GGTACACAGACCAACACTTTCTGCTGTGACCATCTGCTGCATTATTAGGAATAGCTAT	210
OY	196	GTTGATTAACACTAGCAAAAGTCACAGCAAACTTCTTTAAGAAACAGCAACCCCAACTCA	255
Db	211	GTTGATTAACACTAGCAAAAGTCACAGCAAACTTCTTTAAGAAACAGCAACCCCAACTCA	270
OY	256	ACCTATGATACATAT	270
Db	271	ACCTATGATACATAT	285

```

1      RESULT 2
2      US-08-048-164A-3/C
3      Sequence 3, Application US/08048164A
4      Patent No. 5496934
5      GENERAL INFORMATION:
6      APPLICANT: Shoseyov, Oded
7      APPLICANT: Shielegl, Itai
8      APPLICANT: Goldstein, Marc A.
9      APPLICANT: DOL, Roy H.
10     TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
11     NUMBER OF SEQUENCES: 21
12     CORRESPONDENCE ADDRESSES:
13     ADDRESSEE: PENNIE & EDMONDS
14     STREET: 1155 Avenue of the Americas
15     CITY: New York
16     STATE: New York
17     COUNTRY: U.S.A.
18     ZIP: 10036
19     COMPUTER READABLE FORM:
20     MEDIUM TYPE: Floppy disk
21     COMPUTER: IBM PC compatible
22     OPERATING SYSTEM: PC-DOS/MS-DOS
23     SOFTWARE: PatentIn Release #1.0, Version #1.25
24     CURRENT APPLICATION DATA:
25     APPLICATION NUMBER: US/08/048,164A
26     FILING DATE: 14-APR-1993
27     CLASSIFICATION: 435
28     ATTORNEY/AGENT INFORMATION:
29     NAME: Mistock, S. Leslie
30     REGISTRATION NUMBER: 18,872
31     REFERENCE/DOCKET NUMBER: 7809-003
32     TELECOMMUNICATION INFORMATION:
33     TELEPHONE: (212) 790-9090
34     TELEFAX: (212) 869-8864/9741
35     TELEX: 66141 PENNIE
36     INFORMATION FOR SEQ ID NO: 3:
37     SEQUENCE CHARACTERISTICS:
38     LENGTH: 486 base pairs
39     TYPE: nucleic acid
40     STRANDEDNESS: double
41     TOPOLOGY: unknown
42     MOLECULE TYPE: DNA
43     US-08-048-164A-3

```

Query Match	33.28;	Score 255;	DB 1;	Length 486;
Best Local Similarity	100.08;	Pred.No. 5.5e-124;		
Matches 255; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 16 TACAACCTTACAACATGACACACAACAACTCAATTACACCAATTAATCAAAATTACTTAAC 75
|||||
D6 456 TACAACCTTACAACATGACACACAACAACTCAATTACACCAATTAATCAAAATTACTTAAC 397

QY	76	ACATCTGCACGTGATTTAAATTTAAATGACGCTAAAGCTGATGATTTATTTACACAAAGAT	135
Db	396	ACATCTGCACGTGATTTAAATTTAAATGACGCTAAAGCTGATGATTTATTTACACAAAGAT	337
QY	136	GGTACACAGAGACAACCTTCTGGGTGCACATGCTGGTCATATATAGAGAAATAGCTAT	195
Db	336	GGTACACAGAGACAACCTTCTGGGTGCACATGCTGGTCATATATAGAGAAATAGCTAT	277
QY	196	GTTGATTAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACAGCAAGCCCAACATCA	255
Db	276	GTTGATTAACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACAGCAAGCCCAACATCA	217
QY	256	ACCTATGATACATAT 270	
Db	216	ACCTATGATACATAT 202	

```

1      RESULT 3
2      US-08-460-462-1
3      / Sequence 1, Application US/08460462
4      / Patent No. 5670623
5      / GENERAL INFORMATION:
6      / APPLICANT: Shoseyov, Oded
7      / APPLICANT: Shpiegl, Itai
8      / APPLICANT: Goldstein, Marc A.
9      / APPLICANT: Doi, Roy H.
10     / TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
11     / NUMBER OF SEQUENCES: 21
12     / CORRESPONDENCE ADDRESS:
13     / ADDRESSEE: PENNIE & EDMONDS
14     / STREET: 1155 Avenue of the Americas
15     / CITY: New York
16     / STATE: New York
17     / COUNTRY: U.S.A.
18     / ZIP: 10036
19     / COMPUTER READABLE FORM:
20     / MEDIUM TYPE: Floppy disk
21     / COMPUTER: IBM PC compatible
22     / OPERATING SYSTEM: PC-DOS/MS-DOS
23     / SOFTWARE: PatentIn Release #1.0, Version #1.25
24     / CURRENT APPLICATION DATA:
25     / APPLICATION NUMBER: US/08/460,462
26     / FILING DATE: concurrently herewith
27     / CLASSIFICATION: 435
28     / PRIOR APPLICATION DATA:
29     / APPLICATION NUMBER: US 08/048,164
30     / FILING DATE: 14-APR-1993
31     / ATTORNEY/AGENT INFORMATION:
32     / NAME: Mistrock, S. Leslie
33     / REGISTRATION NUMBER: 16,872
34     / REFERENCE/DOCKET NUMBER: 7809-006
35     / TELECOMMUNICATION INFORMATION:
36     / TELEPHONE: (212) 790-9090
37     / TELEFAX: (212) 869-8864/9741
38     / TELEX: 66141 PENNIE
39     / INFORMATION FOR SEQ ID NO: 1:
40     / SEQUENCE CHARACTERISTICS:
41     / LENGTH: 486 base pairs
42     / TYPE: nucleic acid
43     / STRANDEDNESS: double
44     / TOPOLOGY: unknown
45     / MOLECULE TYPE: DNA
46     / FEATURE:
47     / NAME/KEY: CDS
48     / LOCATION: 1..486
49     /
50     / US-08-460-462-1

```

Query Match	33.28;	Score 255;	DB 1;	Length 486;
Best Local Similarity	100.0%;	Pred. No. 5.5e124;		
Matches 255; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

Matches	255;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0
QY	16	TACAACCTTACAATCAGCACAACAACATCATTTACAGCAATATCAAAATTTACTAAC	75						

|||||
Db 31 TACAACCTTAACAANTGACGACAAACAACTCAATTACACCAATTAATCAAAATTACTAAC 90
Qy 76 ACATCTGACGAGTATTAATTAATGAGTAAAGTTGATATTTACACAGGAT 135
Db 91 ACATCTGACGAGTATTAATTAATGAGTAAAGTTGATATTTACACAGGAT 150
Qy 136 GTTACACAAAGACAACTTCTGCTGACCAATGCTGTCATTTATAGAAATAGCTAT 195
Db 151 GTTACACAAAGACAACTTCTGCTGACCAATGCTGTCATTTATAGAAATAGCTAT 210
Qy 196 GTTGATTAACACTAGCAAAAGTGACGACAACTTGCTTAAGAAACAGCAGCCCAATCA 255
Db 211 GTTGATTAACACTAGCAAAAGTGACGACAACTTGCTTAAGAAACAGCAGCCCAATCA 270
Qy 256 ACCTATGATACATAT 270
Db 271 ACCTATGATACATAT 285

RESULT 4

US-08-460-462-3/C
Sequence 3, Application US/08460462
Patent No. 5670623
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shlegel, Itai
APPLICANT: Goldstein, Marc A.
APPLICANT: Dol, Roy H.
TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,462
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-08-460-462-3

Query Match 33.2%; Score 255; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 5,5e-124;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 16 TACAACCTTAACAANTGACGACAAACAACTCAATTACACCAATTAATCAAAATTACTAAC 75
|||||

Db 456 TACAACCTTAACAANTGACGACAAACAACTCAATTACACCAATTAATCAAAATTACTAAC 397
Qy 76 ACATCTGACGAGTATTAATTAATGAGTAAAGTTGATATTTACACAGGAT 135
Db 396 ACATCTGACGAGTATTAATTAATGAGTAAAGTTGATATTTACACAGGAT 337
Qy 136 GTTACACAAAGACAACTTCTGCTGACCAATGCTGTCATTTATAGAAATAGCTAT 195
Db 336 GTTACACAAAGACAACTTCTGCTGACCAATGCTGTCATTTATAGAAATAGCTAT 277
Qy 196 GTTGATTAACACTAGCAAAAGTGACGACAACTTGCTTAAGAAACAGCAGCCCAATCA 255
Db 276 GTTGATTAACACTAGCAAAAGTGACGACAACTTGCTTAAGAAACAGCAGCCCAATCA 217
Qy 256 ACCTATGATACATAT 270
Db 216 ACCTATGATACATAT 202

RESULT 5

US-08-460-457-1
Sequence 1, Application US/08460457
Patent No. 5719044
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shlegel, Itai
APPLICANT: Goldstein, Marc A.
APPLICANT: Dol, Roy H.
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN FUSION PROTEINS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,457
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..486
US-08-460-457-1

Query Match 33.2%; Score 255; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 5,5e-124;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 31 TACAACTCTAACAAATCAGCACAACAACTCAATTACCAATTAATCAAAATTACTAAC 90
QY 76 ACATCTGACAGTGTATTAATTAATGAGTAAAGTTAGATATTATTACAGAGTAT 135
Db 91 ACATCTGACAGTGTATTAATTAATGAGTAAAGTTAGATATTATTACAGAGTAT 150
QY 136 GGTACACAAAGGACAACTTCTGTGTGACCATCTGTGTCATTTATAGGAATAGCTAT 195
Db 151 GGTACACAAAGGACAACTTCTGTGTGACCATCTGTGTCATTTATAGGAATAGCTAT 210
QY 196 GTTGATTAACACTAGCAAAAGTGACAGCAAACTGCTTAAGAAACAGCAAGCCCAACATCA 255
Db 211 GTTGATTAACACTAGCAAAAGTGACAGCAAACTGCTTAAGAAACAGCAAGCCCAACATCA 270
QY 256 ACCTATGATACATAT 270
Db 271 ACCTATGATACATAT 285

RESULT 8
US-08-460-458-3/C
Sequence 3, Application US/08460458
Patent No. 5738984
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
TITLE OF INVENTION: KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAIN
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,458
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-08-460-458-3

Query Match 33.2%; Score 255; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 5,5e-124;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGACGACAAACAACCTCAATTACACCAATTAATCAAAATTACTAAC 75
Db 456 TACAACCTTAACAATGACGACAAACAACCTCAATTACACCAATTAATCAAAATTACTAAC 397
QY 76 ACATCTGACAGTGTATTAATTAATGAGTAAAGTTAGATATTATTACAGAGTAT 135

Db 396 ACATCTGACAGTGTATTAATTAATGAGTAAAGTTAGATATTATTACAGAGTAT 337
QY 136 GGTACACAAAGGACAACTTCTGTGTGACCATCTGTGTCATTTATAGGAATAGCTAT 195
Db 336 GGTACACAAAGGACAACTTCTGTGTGACCATCTGTGTCATTTATAGGAATAGCTAT 277
QY 196 GTTGATTAACACTAGCAAAAGTGACAGCAAACTGCTTAAGAAACAGCAAGCCCAACATCA 255
Db 276 GTTGATTAACACTAGCAAAAGTGACAGCAAACTGCTTAAGAAACAGCAAGCCCAACATCA 217
QY 256 ACCTATGATACATAT 270
Db 216 ACCTATGATACATAT 202

RESULT 9
US-08-460-455-1
Sequence 1, Application US/08460455
Patent No. 5837814
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shiple, Ital
APPLICANT: Goldstein, Marc A.
APPLICANT: Doi, Roy H.
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN PROTEINS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,455
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..486
US-08-460-455-1

Query Match 33.2%; Score 255; DB 2; Length 486;
Best Local Similarity 100.0%; Pred. No. 5,5e-124;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGACGACAAACAACCTCAATTACACCAATTAATCAAAATTACTAAC 75
Db 31 TACAACCTTAACAATGACGACAAACAACCTCAATTACACCAATTAATCAAAATTACTAAC 90

OY	76	ACATCTGCAGTGGATTAAATTTAAATGACGTAAAGAATTAGATTTATTACACAAGGAT	135
Db	91	ACATCTGCAGTGGATTAAATTTAAATGACGTAAAGAATTAGATTTATTACACAAGGAT	150
OY	136	GSTACACAAGGCAACAACCTTTGGTGTCACCATCGTGCATATATAGAAAATAGCTAT	195
Db	151	GSTACACAAGGCAACAACCTTTGGTGTCACCATCGTGCATATATAGGAATAGCTAT	210
OY	196	GTTGATTAACACTAGCAAGAAGTACACGCAAACTTTCGTTAAAGCAACGCAAGCCCAACATCA	255
Db	211	GTTGATTAACACTAGCAAGAAGTACACGCAAACTTTCGTTAAAGCAACGCAAGCCCAACATCA	270
OY	256	ACCTATGATACATAT	270
Db	271	ACCTATGATACATAT	285

```

1      RESULT 10
2      US-08-460-455-3/C
3      Sequence 3, Application US/08460455
4      Patent No. 5837814
5      GENERAL INFORMATION:
6      APPLICANT: Shoseyov, Oded
7      APPLICANT: Shoseyov, Itai
8      APPLICANT: Goldstein, Marc A.
9      APPLICANT: Doi, Roy H.
10     TITLE OF INVENTION: CELLULOSE BINDING DOMAIN PRO
11     NUMBER OF SEQUENCES: 21
12     CORRESPONDENCE ADDRESS:
13     ADDRESSEE: PENNIE & EDMONDS
14     STREET: 1155 Avenue of the Americas
15     CITY: New York
16     STATE: New York
17     COUNTRY: U.S.A.
18     ZIP: 10036
19
20     COMPUTER READABLE FORM:
21     MEDIUM TYPE: floppy disk
22     COMPUTER: IBM PC compatible
23     OPERATING SYSTEM: PC-DOS/MS-DOS
24     SOFTWARE: PatentIn Release #1.0, Version #1.25
25     CURRENT APPLICATION DATA:
26     APPLICATION NUMBER: US/08/460,455
27     FILING DATE: currently herewith
28     CLASSIFICATION: 435
29
30     PRIOR APPLICATION DATA:
31     APPLICATION NUMBER: US 08/048,164
32     FILING DATE: 14-APR-1993
33     ATTORNEY/AGENT INFORMATION:
34     NAME: Mistrock, S. Leslie
35     REGISTRATION NUMBER: 18,872
36     REFERENCE/DOCKET NUMBER: 7809-009
37     TELECOMMUNICATION INFORMATION:
38     TELEPHONE: (212) 790-9090
39     TELEFAX: (212) 869-8864/9741
40     TELE: 66141 PENNIE
41
42     INFORMATION FOR SEQ ID NO: 3:
43     SEQUENCE CHARACTERISTICS:
44     LENGTH: 486 base pairs
45     TYPE: nucleic acid
46     STRANDEDNESS: double
47     TOPOLOGY: unknown
48
49     MOLECULE TYPE: DNA
50     US-08-460-455-3

```

Query Match	33.28;	Score 255;	DB 2;	Length 486;
Best Local Similarity	100.0%;	Pred. No. 5.5e-124;		
Matches 255; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

Db	396	ACATGTGCAGTGAATTAAATTAAATGAGCTAAAGTTGATATTATTACACAGTGAT	337
Qy	136	GGTACACAGACACAACACTTTTGTGTGACCATCTGTCATTATTAGGAATAGCTAT	195
Db	336	GGTAAACACAGACAAAACCTTTCTGTGTGCACCATCTGTGATTAATTAGGAATAGCTAT	277
Qy	136	GTTGATTACACTAGCAAAAGTACACGACAACTTCTTTAAGAAACAGACGCCAACATCA	255
Db	276	GTTGATTACACTAGCAAAAGTACACGACAACTTCTTTAAGAAACAGACGCCAACATCA	217
Qy	256	ACCTATGATTACATAAT	270
Db	216	ACCTATGATTACATAAT	202

```

1  RESULT 11
2  US-08-330-394A-1
3  Sequence 1, Application US/08330394A
4  Patent No. 5856201
5  GENERAL INFORMATION:
6  APPLICANT: Shoseyov, Oded
7  APPLICANT: yosef, karmey
8  APPLICANT: Shpiegl, Itai
9  APPLICANT: Goldstein, Marc A.
10 APPLICANT: Dol, Roy H.
11 TITLE OF INVENTION: METHODS OF DETECTION USING THE
12 TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
13 NUMBER OF SEQUENCES: 33
14 CORRESPONDENCE ADDRESS:
15 ADDRESSEE: PENNIE & EDMONDS
16 STREET: 115 AVENUE OF THE AMERICAS
17 CITY: NEW YORK
18 STATE: NEW YORK
19 COUNTRY: U.S.A.
20 ZIP: 10036
21 COMPUTER READABLE FORM:
22 MEDIUM TYPE: Floppy disk
23 COMPUTER: IBM PC compatible
24 OPERATING SYSTEM: PC-DOS/MS-DOS
25 SOFTWARE: PatentIn Release #1.0, Version #1.30
26 CURRENT APPLICATION DATA:
27 APPLICATION NUMBER: US/08/330,394A
28 FILING DATE: 27-OCT-1994
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: CIP OF PCT/US94/04132
31 FILING DATE: 14-APR-1994
32 ATTORNEY/AGENT INFORMATION:
33 NAME: MISROCK, S. LESLIE
34 REFERENCE/DOCKET NUMBER: 7809-005
35 TELECOMMUNICATION INFORMATION:
36 TELEPHONE: (212) 790-9090
37 TELEFAX: (212) 869-8864
38 TELEX: 66441 PENNIE
39 INFORMATION FOR SEQ ID NO: 1:
40 SEQUENCE CHARACTERISTICS:
41 LENGTH: 486 base pairs
42 TYPE: nucleic acid
43 STRANDEDNESS: double
44 TOPOLOGY: unknown
45 MOLECULE TYPE: DNA
46 FEATURE:
47 NAME/KEY: CDS
48 LOCATION: 1..486
49 US-08-330-394A-1

```

Query Match	33.28;	Score 255;	DB 2;	Length 486;
Best Local Similarity	100.0%;	Pred. No.	5.5e-124;	
Matches 255: Conservative	0;	Mismatches	0;	Indels 0;
				Gaps 0;

OY 16 TACAACCTTACAATCAGCACAAACAACCTCAATTACACCAATTAATCAAATTTACTAAC 75
|||||
Db 31 TACAACCTTACAATCAGCACAAACAACCTCAATTACACCAATTAATCAAATTTACTAAC 90

Oy 16 TACAACCTCTAACCAAAATCAGCACAAAACCAACTCAATTACACCAATATCAAATTACTTAAC 75
|||||
Db 456 TACAACCTCTAACCAAAATCAGCACAAAACCAACTCAATTACACCAATATCAAATTACTTAAC 397

Db 96 ACATCTGCACAGTGAATTTAAATTTAAAGTACGTAAAGTTAGATTTATTACACAAAGTAT 155

|||||
Db 156 GGTACACAGAGCAAACTTCTGGTGACCATCTGTGCATTATTAGGAATAGCTAT 215
QY 196 GTTGATTAACACTAGACAGAGCAAACTGCTTAAGAAACAGAGCCCAACTCA 255
Db 216 GTTGATTAACACTAGACAGAGCAAACTGCTTAAGAAACAGAGCCCAACTCA 275
QY 256 ACCATGATACATAT 270
Db 276 ACCATGATACATAT 290

RESULT 14
US-09-006-632-5

; Sequence 5, Application US/09006632
; Patent No. 6184440
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shani, Ziv
; APPLICANT: Shpigel, Etai
; TITLE OF INVENTION: TRANSGENIC PLANTS OF ALTERED
; MORPHOLOGY
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/006,632
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7809-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 499 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-09-006-632-5

Query Match 33.2%; Score 255; DB 4; Length 499;
Best Local Similarity 100.0%; Pred. No. 5.5e-124;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTCTAACAATGAGCAACAACCAATCAATTACACCAATTAATCAAAATTACTAAC 75
Db 36 TACAACCTCTAACAATGAGCAACAACCAATCAATTACACCAATTAATCAAAATTACTAAC 95
QY 76 ACATCTGACAGTGAATTTAATTAATGAGGTAAAGTTAGATATTATTACACAAGTGAT 135
Db 96 ACATCTGACAGTGAATTTAATTAATGAGGTAAAGTTAGATATTATTACACAAGTGAT 155
QY 136 GGTACACAGAGCAAACTTCTGGTGACCATCTGTGCATTATTAGGAATAGCTAT 195
Db 156 GGTACACAGAGCAAACTTCTGGTGACCATCTGTGCATTATTAGGAATAGCTAT 215
QY 196 GTTGATTAACACTAGACAGAGCAAACTGCTTAAGAAACAGAGCCCAACTCA 255

|||||
Db 216 GTTGATTAACACTAGACAGAGCAAACTGCTTAAGAAACAGAGCCCAACTCA 275
QY 256 ACCATGATACATAT 270
Db 276 ACCATGATACATAT 290

RESULT 15
US-09-277-716-21

; Sequence 21, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Artificial Sequence: fusion construct
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1146)
; FEATURE:
; OTHER INFORMATION: DNA encoding a Cellulose Binding Domain (CBD)-Gaussia lucifera
; US-09-277-716-21

Query Match 33.2%; Score 255; DB 4; Length 1146;
Best Local Similarity 100.0%; Pred. No. 5.5e-124;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTCTAACAATGAGCAACAACCAATCAATTACACCAATTAATCAAAATTACTAAC 75
Db 16 TACAACCTCTAACAATGAGCAACAACCAATCAATTACACCAATTAATCAAAATTACTAAC 75
QY 76 ACATCTGACAGTGAATTTAATTAATGAGGTAAAGTTAGATATTATTACACAAGTGAT 135
Db 76 ACATCTGACAGTGAATTTAATTAATGAGGTAAAGTTAGATATTATTACACAAGTGAT 135
QY 136 GGTACACAGAGCAAACTTCTGGTGACCATCTGTGCATTATTAGGAATAGCTAT 195
Db 136 GGTACACAGAGCAAACTTCTGGTGACCATCTGTGCATTATTAGGAATAGCTAT 195
QY 196 GTTGATTAACACTAGACAGAGCAAACTGCTTAAGAAACAGAGCCCAACTCA 255
Db 196 GTTGATTAACACTAGACAGAGCAAACTGCTTAAGAAACAGAGCCCAACTCA 255
QY 256 ACCATGATACATAT 270
Db 256 ACCATGATACATAT 270

Search completed: October 27, 2002, 23:22:28
Job time : 63.0363 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 17:46:54 ; Search time 194.512 Seconds
(without alignments)
6777.910 Million cell updates/sec

Title: US-09-677-374-5

Sequence: 1 atgtcagttgaattctacaa.....agtgattagcaccgaaaaa 768

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

word size :

Total number of hits satisfying chosen parameters: 3472872

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*

1:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	768	100.0	768	22	AAF86248	DNA sequence of c
2	483	62.9	483	22	AAF86247	DNA sequence of E
3	270	35.2	573	22	AAID1043	Clostridium cellul
4	270	35.2	1030	22	AAID1044	Clostridium cellul
5	255	33.2	486	15	AAO72917	cellulose binding
6	255	33.2	486	20	AAV74072	C. cellulovorans C
7	255	33.2	499	20	AAK24930	Clostridium cellul
8	255	33.2	507	22	AAID1042	Clostridium cellul
9	255	33.2	584	22	AAID1046	Chimeric S peptide

10	255	33.2	1146	20	AAZ27550	Gaussia luciferase
11	255	33.2	1146	24	AAZ22201	Gaussia species CB
12	255	33.2	1288	22	AAD11045	Clostridium celluli
13	118	15.4	118	22	AAF66252	PCR primer #4 used
14	104	13.5	110	22	AAF66254	PCR primer #2 used
15	102	13.3	102	22	AAF66251	PCR primer #5 used
16	87	11.3	111	22	AAF66255	PCR primer #1 used
17	75	9.8	94	22	AAF66253	PCR primer #3 used
18	40	5.2	110	22	AAF66256	PCR primer #6 used
19	34	4.4	38	21	AAA27285	PCR primer for tru
20	20	2.6	26	20	AAV74089	Plasmid pET-R1-CBD
21	20	2.6	3833	23	ABL02577	Drosophila melanog
22	19	2.5	808	19	AAV28688	Ripening banana pu
23	19	2.5	892	22	AAH07401	Human cDNA clone (
24	19	2.5	1438	21	AAZ45336	DNA encoding a man
25	19	2.5	1482	20	AAK90978	DNA encoding pecta
26	19	2.5	1482	20	AAZ31562	pectate lyase CBD
27	27	2.5	2042	22	AAH16456	Human cDNA sequenc
28	19	2.5	2562	22	AAI59154	Human polyomucleo
29	19	2.5	2735	22	AAF32744	Human secreted pro
30	19	2.5	2878	22	AAK02416	Human secreted pro
31	19	2.5	4283	22	AAK25994	Human cDNA encodin
32	19	2.5	5562	19	AAT86625	C. thermocellum C1
33	19	2.5	6223	22	AAI58572	Human polyomucleo
34	19	2.5	6304	22	AAI58574	Human polyomucleo
35	19	2.5	6382	22	AAI58573	Human polyomucleo
36	18	2.3	1029	22	AAH52010	Mycobacterium tube
37	18	2.3	1237	22	AAK94517	DNA encoding novel
38	18	2.3	1300	18	AAT66951	Asiaglycoprotein
39	18	2.3	1353	21	AAC39312	Arabidopsis thalia
40	18	2.3	1356	23	AAK559780	Propionibacterium
41	18	2.3	1445	23	AAK77372	DNA encoding novel
42	18	2.3	1623	23	AAK82377	DNA encoding novel
43	18	2.3	1903	22	AAD16584	Human novel protei
44	18	2.3	3684	22	AAK64335	Human novel protei
45	18	2.3	3741	22	AAH81357	DNA encoding novel
					Escherichia coli p	

ALIGNMENTS

RESULT	1
AAAF6248	
ID	AAF6248 standard; DNA; 768 bp.
XX	
AC	AAF6248;
XX	
DT	11-JUL-2001 (first entry)
XX	
DE	DNA sequence of c17e2 ospA construct with N-terminal fusion partner.
XX	
KW	Polyketide synthase, Pseudomonas fluorescens Pf-0; rickettsial pathogen;
KW	vaccine; OspA; salmonella typhimurium DT104; rickettsial disease;
SR	SRS; 17E2; fusion construct; ds.
XX	
OS	Piscirickettsia salmonis.
OS	Synthetic.
XX	
FH	Key
FT	CDS
XX	
FT	Location/Qualifiers
FT	1..768
FT	/tag= a
FT	/partial
FT	/product= "C17E2 OspA with N-terminal fusion partner"
FT	/note= "No stop codon is given"
FT	1..285
FT	/tag= b
FT	/note= "DNA encoding undefined N-terminal fusion partner"
FT	286..768
FT	/tag= C
FT	/note= "Optimised OspA construct c17E2"
XX	
CA2281913-A1.	

PD 17-MAR-2001.
 XX
 PF 17-SEP-1999; 99CA-2281913.
 XX
 PR 17-SEP-1999; 99CA-2281913.
 XX
 PA (KAYW/) KAY W W.
 PA (BURI/) BURIAN J.
 PA (KUZU/) KUZUK M A.
 XX
 PI Kay WW, Burian J, Kuzuk MA.
 XX
 DR WPI: 2001-316844/34.
 DR P-PSDB; AAB81128.
 PT Method for protecting poikilothermic fish against salmonid rickettsial
 PT septicemia and other rickettsial diseases comprises administering a
 PT vaccine containing the Ospa protein of *Piscirickettsia salmonis*.
 XX
 PS Example 4; Fig 5; 35pp; English.
 XX
 CC This invention relates to a method for the protection against infection
 CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*
 CC *salmonis*. The method comprises administering an immunogenic amount of a
 CC *P. salmonis* specific antigen termed Ospa, or an immunogenic fragment of
 CC Ospa in the form of a vaccine. The method is used for protecting animals,
 CC particularly poikilothermic fish, against the bacterial pathogen
 CC *P. salmonis*. The method is also useful for protecting against salmonid
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
 CC sequence represents *P. salmonis* Ospa DNA termed C17E2 optimised for
 CC expression in *Escherichia coli* fused to DNA encoding an undefined
 CC N-terminal fusion partner. The protein encoded by this fusion construct
 CC is used in a vaccine to create an anti-Ospa antibody response.
 XX
 SQ Sequence 768 BP; 220 A; 183 C; 190 G; 175 T; 0 other:

Query Match 100.0%; Score 768; DB 22; Length 768;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 768; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAATTAATTTCTAACAATCTAACAATCAGACAAACAACTAATTAACCAATA 60
 DB 1 ATGCAATTAATTTCTAACAATCTAACAATCAGACAAACAACTAATTAACCAATA 60
 QY 61 ATGCAATTAATTTCTAACAATCTAACAATCAGACAAACAACTAATTAACCAATA 120
 DB 61 ATGCAATTAATTTCTAACAATCTAACAATCAGACAAACAACTAATTAACCAATA 120
 QY 121 TATTACAAGTAGTAGTACACAAGACAACCTTTCTGTTGTCACATGCTGTCATTA 180
 DB 121 TATTACAAGTAGTAGTACACAAGACAACCTTTCTGTTGTCACATGCTGTCATTA 180
 QY 181 TTAGGAATAGCTATGTTGATACACTAGCAAGTGCAGCAACTTCGTTAAAGAACCA 240
 DB 181 TTAGGAATAGCTATGTTGATACACTAGCAAGTGCAGCAACTTCGTTAAAGAACCA 240
 QY 241 GCAAGCCCAACATCAACCTATGATACATCTGATCCGTCCTATATGCGTGGTTCG 300
 DB 241 GCAAGCCCAACATCAACCTATGATACATCTGATCCGTCCTATATGCGTGGTTCG 300
 QY 301 CAGGCGAGCTCTCTGATCATATATCTGTTCTGTTGTTGCGGCCAGAACTTACG 360
 DB 301 CAGGCGAGCTCTCTGATCATATATCTGTTCTGTTGTTGCGGCCAGAACTTACG 360
 QY 361 CGCCAGCAAGTTGGCGCGCCACCGGTCGTTGCGGCTGTTGCGGCCAGCTTTC 420
 DB 361 CGCCAGCAAGTTGGCGCGCCACCGGTCGTTGCGGCTGTTGCGGCCAGCTTTC 420
 QY 421 GGTAAAGCTCTGCTGTCGATGGCATGGCGGTCGGGTCGGGCGCTCATTT 480
 DB 421 GGTAAAGCTCTGCTGTCGATGGCATGGCGGTCGGGTCGGGCGCTCATTT 480
 QY 481 GGCTCTAAATCGCTCAGACATGACAGCAGAGATAAATCAAACTGAACCAAGTCTCTG 540

DB 481 GGCTCTAAATCGCTCAGACATGACAGCAGAGATAAATCAAACTGAACCAAGTCTCTG 540
 QY 541 GAAAGAGTAAAGCGCGCAGGTTACTGCTTGGCTATATCCGACACCGGTTAACAGCTAC 600
 DB 541 GAAAGAGTAAAGCGCGCAGGTTACTGCTTGGCTATATCCGACACCGGTTAACAGCTAC 600
 QY 601 TCTGTGGAACCGGTTGCGACCTACACCGGTTACACAAACAGGACCGCTCAGCACTAC 660
 DB 601 TCTGTGGAACCGGTTGCGACCTACACCGGTTACACAAACAGGACCGCTCAGCACTAC 660
 QY 661 TCGCGCAATTTTACAGCAAGAACCCATGATCGACAGTTCAGAAACAGAAATCTACGCGACC 720
 DB 661 TCGCGCAATTTTACAGCAAGAACCCATGATCGACAGTTCAGAAACAGAAATCTACGCGACC 720
 QY 721 GCGTGCCTCAGCCGATGCGCGCTGCGAGGTATTTGACACCGCAAAA 768
 DB 721 GCGTGCCTCAGCCGATGCGCGCTGCGAGGTATTTGACACCGCAAAA 768

RESULT 2

AAFB6247
 ID AAFB6247 standard; DNA; 483 BP.
 XX
 AC AAFB6247;
 XX
 DT 11-JUL-2001 (first entry)
 XX
 DE DNA sequence of *E. coli* optimised ospa gene 17E2.
 XX
 KW Poikilothermic fish; *Piscirickettsia salmonis*; rickettsial pathogen;
 KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;
 KW SRS; 17E2; ds.
 XX
 OS *Piscirickettsia salmonis*.
 OS Synthetic.
 OS
 FH Key Location/Qualifiers
 FT 1..483
 FT CDS
 FT /tag= a
 FT /partial
 FT /product= "Ospa"
 FT /note= "Genus specific 17kDa antigen, the sequence does
 not include a stop codon"
 XX
 PN CA281913-A1.
 XX
 PD 17-MAR-2001.
 XX
 PF 17-SEP-1999; 99CA-2281913.
 XX
 PR 17-SEP-1999; 99CA-2281913.
 XX
 PA (KAYW/) KAY W W.
 PA (BURI/) BURIAN J.
 PA (KUZU/) KUZUK M A.
 XX
 PI Kay WW, Burian J, Kuzuk MA.
 XX
 DR WPI: 2001-316844/34.
 DR P-PSDB; AAB81128.
 PT Method for protecting poikilothermic fish against salmonid rickettsial
 PT septicemia and other rickettsial diseases comprises administering a
 PT vaccine containing the Ospa protein of *Piscirickettsia salmonis*.
 XX
 PS Example 3; Fig 4C; 35pp; English.
 XX
 CC This invention relates to a method for the protection against infection
 CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*
 CC *salmonis*. The method comprises administering an immunogenic amount of a
 CC *P. salmonis* specific antigen termed Ospa, or an immunogenic fragment of
 CC Ospa in the form of a vaccine. The method is used for protecting animals,

CC particularly polikilothermic fish, against the bacterial pathogen
 CC P. salmonis. The method is also useful for protecting against salmonid
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
 CC sequence represents DNA which has been optimised for expression in
 CC Escherichia coli to encode the P. salmonis OspA protein. An OspA protein
 CC with an N-terminal fusion partner is used in a vaccine to create an
 CC anti-OspA antibody response.

SO Sequence 483 BP: 108 A; 127 C; 148 G; 100 T; 0 other;

Query Match

Best Local Similarity 62.9%; Score 483; DB 22; Length 483;
 Pred. No. 6, 1e-242;

Matches 483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 286 ATGCGTGTGCTCTGCGAGGCGAGCTCTGTATATATATCTGTGTTTCCGTGGGTGC 345
 DB 1 ATGCGGTGTGCTCTGCGAGGCGAGCTCTGTATATATCTGTGTTTCCGTGGGTGC 60
 OY 346 GCCCAGAACTTCAAGCCCGAGAAATTGGCGGCCACCGGTGGTGTGGCGGTGTT 405
 DB 61 GCCCAGAACTTCAAGCCCGAGAAATTGGCGGCCACCGGTGGTGTGGCGGTGTT 120
 OY 406 GCCGCGCACTGTTCGTAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 465
 DB 121 GCCGCGCACTGTTCGTAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180
 OY 466 CTGGGCGGTGTGATGCTCTTAAATCGGTCAAGCATGACACGAGATTAATATAA 525
 DB 181 CTGGGCGGTGTGATGCTCTTAAATCGGTCAAGCATGACACGAGATTAATATAA 240
 OY 526 CTGACACAGTCTGTGAAAAAGTAAAGCGGCGGAGTTCTGTTGGGTAAATCCGGAC 585
 DB 241 CTGACACAGTCTGTGAAAAAGTAAAGCGGCGGAGTTCTGTTGGGTAAATCCGGAC 300
 OY 586 ACCGCTACAGCTACTGTGTGAACCGGTTCCGACCTACAGCGTTTACAAACAGAGAA 645
 DB 301 ACCGCTACAGCTACTGTGTGAACCGGTTCCGACCTACAGCGTTTACAAACAGAGAA 360
 OY 646 CGCGCTACAGCTACTGTGTGAACCGGTTCCGACCTACAGCGTTTACAAACAGAGAA 705
 DB 361 CGCGCTACAGCTACTGTGTGAACCGGTTCCGACCTACAGCGTTTACAAACAGAGAA 420
 OY 706 GAATCTACAGCAGCGGCTCCCTCAGCCGATGCGGTGAGGTATGACCGGAA 765
 DB 421 GAATCTACAGCAGCGGCTCCCTCAGCCGATGCGGTGAGGTATGACCGGAA 480
 OY 766 AAA 768
 DB 481 AAA 483

RESULT 3

AAD11043 ID AAD11043 standard; DNA: 573 BP.

AC AAD11043;

DT 24-SEP-2001 (first entry)

DE Clostridium cellulovorans cellulose binding domain-180 (CBD-180) DNA.

XX polysaccharide modification; polysaccharide binding domain; PBD; paper;

KW yarn; fiber; textile; biological crosslinker; mechanical property;

KW Clostridium cellulovorans cellulose binding domain; wet strength;

KW durability; elasticity; CBDClos; cellulose binding protein A; CBP A;

XX CBD-180; ds.

OS Clostridium cellulovorans.

PH key Location/Qualifiers
 FT CDS 3..572
 FT /tag- a
 FT /product- "Cellulose binding domain-180 protein"

FT /note= "CDS does not include stop codon"
 FT /partial

PN WO200134091-A2.

PD 17-MAY-2001.

PF 02-NOV-2000; 2000WO-IL00708.

PR 08-NOV-1999; 99US-0164140.

PR 18-NOV-1999; 99US-0166389.

PA (CBDF-) CBD TECHNOLOGIES LTD.

PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.

PI Levy I, Nussinovitch A, Shoseyov O;

DR WPI: 2001-457121/49.

DR P-PsDB; AAE05746.

PT Preparation of a polysaccharide containing material having at least one

PT desired structural, chemical, physical, electrical and/or mechanical

PT property -

PS Example 1.2; Fig 1e-1g; 121pp; English.

CC The present invention relates to methods and compositions for cross-
 CC linking and/or modifying the properties of polysaccharide materials.
 CC The method involves treating the polysaccharide structure with a
 CC polysaccharide binding domain (PBD) fusion protein. The method is
 CC used to alter the structural, chemical, physical, electrical and
 CC mechanical properties of polysaccharide materials such as paper,
 CC yarns, fibers and textiles, using biological crosslinking agents.
 CC The polysaccharide containing materials have improved mechanical
 CC properties such as wet strengths, durability and elasticity. The PBD
 CC reagent is applied in the forming stage in futing paper manufacture
 CC which eliminates the sizing step. The use of a biological crosslinker
 CC improves the recyclability of paper products. The PBD reagent maintains
 CC the fine fibers in a slurry therefore resulting in better recovery of
 CC raw materials. The PBD molecules are eluted by strong alkaline conditions
 CC which enhances the ability of the alkaline glue used in binding to
 CC penetrate paper.
 CC The present sequence is a DNA encoding Clostridium cellulovorans
 CC cellulose binding domain (CBD)-180. CBD-180 contains 180 amino acids of
 CC Clostridium cellulovorans CBD (CBDClos) of cellulose binding protein A
 CC (CBP A).

SO Sequence 573 BP; 215 A; 112 C; 98 G; 148 T; 0 other;

Query Match 35.2%; Score 270; DB 22; Length 573;

Best Local Similarity 100.0%; Pred. No. 1e-130;

Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGTCAGTTGAATCTTACAACTCTAACAAATCGACCAACAAACCTCAATTACACCAATA 60
 DB 3 ATGTCAGTTGAATCTTACAACTCTAACAAATCGACCAACAAACCTCAATTACACCAATA 62
 OY 61 ATCAAAATTTACTACACATGTGACAGTATTAAATTTAAATGACGTAAGTTAGTAT 120
 DB 63 ATCAAAATTTACTACACATGTGACAGTATTAAATTTAAATGACGTAAGTTAGTAT 122
 OY 121 TATTACACAGTATGTGTACAAAGCAAACTTTCGTGTGATGACCAATGCTGTGCACTTA 180
 DB 123 TATTACACAGTATGTGTACAAAGCAAACTTTCGTGTGATGACCAATGCTGTGCACTTA 182
 OY 181 TTAGGAATATAGTATGTGTGTAACACTAGCAAGTGACAGCAAACTTCGTTAAGAACA 240
 DB 183 TTAGGAATATAGTATGTGTGTAACACTAGCAAGTGACAGCAAACTTCGTTAAGAACA 242
 OY 241 GCAAGCCCAACATCACTTATGATACATAT 270
 DB 243 GCAAGCCCAACATCACTTATGATACATAT 272

```
RESULT 4
AADI1044
ID AADI1044 standard; DNA: 1030 BP.
XX
AC AADI1044;
XX
DT 24-SEP-2001 (first entry)
XX
DE Clostridium cellulovorans CBD cross linker protein (CCP) encoding DNA.
XX
KW Polysaccharide modification; polysaccharide binding domain; PBD; paper;
KW yarn; fiber; textile; biological crosslinker; mechanical property;
KW web strength; durability; elasticity; cellulose binding domain; CBD;
XX
OS Clostridium cellulovorans.
XX
FH Key Location/Qualifiers
FT CDS 3..1028
FT /*tag= a
FT /product= "CCP protein"
FT /transl_except= "(pos: 1020..1028, aa:Pro-Asp)"
FT /note= "CDS does not include stop codon"
FT /partial
XX
PN WO200134091-A2.
XX
PD 17-MAY-2001.
XX
PF 02-NOV-2000; 2000WO-IL00708.
XX
PR 08-NOV-1999; 99US-0164140.
PR 18-NOV-1999; 99US-0166389.
XX
PA (CBD-) CBD TECHNOLOGIES LTD.
PA (YISS ) YISSUM RES DEV CO HERREW UNIV JERUSALEM.
XX
PI Levy I, Nussinovitch A, Shoseyov O;
XX
DR MPI: 2001-457121/49.
DR P-PSDB; AAE05747.
XX
PT Preparation of a polysaccharide containing material having at least one
PT desired structural, chemical, physical, electrical and/or mechanical
PT property
XX
PS Example 1.2; Fig 2b-2e; 121pp; English.
XX
CC The present invention relates to methods and compositions for cross-
CC linking and/or modifying the properties of polysaccharide materials.
CC The method involves treating the polysaccharide structure with a
CC polysaccharide binding domain (PBD) fusion protein. The method is
CC used to alter the structural, chemical, physical, electrical and
CC mechanical properties of polysaccharide materials such as paper,
CC yarns, fibers and textiles, using biological crosslinking agents.
CC The polysaccharide containing materials have improved mechanical
CC properties such as wet strengths, durability and elasticity. The PBD
CC reagent is applied in the forming stage in futing, paper manufacture
CC which eliminates the sizing step. The use of a biological crosslinker
CC improves the recyclability of paper products. The PBD reagent maintains
CC the fine fibers in a slurry therefore resulting in better recovery of
CC raw materials. The PBD molecules are eluted by strong alkaline conditions
CC which enhances the ability of the alkaline glue used in binding to
CC penetrate paper.
CC The present sequence is a DNA encoding Clostridium cellulovorans
CC cellulose binding domain (CBD) cross linker protein (CCP).
XX
SO Sequence 1030 BP; 402 A; 195 C; 166 G; 267 T; 0 other;
Query Match 35.2%; Score 270; DB 22; Length 1030;
Best Local Similarity 100.0%; Pred. No. 1e-130;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ATGTCAGTTGAATCTCTACAGCTCTACAAATCAGCAGCAACAACTCAATTACACCATATA 60
DB 3 ATGTCAGTTGAATCTCTACAGCTCTACAAATCAGCAGCAACAACTCAATTACACCATATA 62
QY 61 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATTTAAATGAGTAAAGTTAGATAT 120
DB 63 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATTTAAATGAGTAAAGTTAGATAT 122
QY 121 TATTACACAGTGTGATGACACAGCAACAACTTTGCTGTGACCAATGCTGTGATTA 180
DB 123 TATTACACAGTGTGATGACACAGCAACAACTTTGCTGTGACCAATGCTGTGATTA 182
QY 181 TTAGGAATTTAGCTATGTTGATTAACACTAGCAAAAGTGACAGCAAACTGCTTAAGAACA 240
DB 183 TTAGGAATTTAGCTATGTTGATTAACACTAGCAAAAGTGACAGCAAACTGCTTAAGAACA 242
QY 241 GCAAGCCCAACATCAACCTATGATACATAT 270
DB 243 GCAAGCCCAACATCAACCTATGATACATAT 272
```

```
RESULT 5
AAQ72917
ID AAQ72917 standard; DNA: 486 BP.
XX
AC AAQ72917;
XX
DT 05-JUN-1995 (first entry)
XX
DE Cellulose binding domain.
XX
KW Cellulose binding domain; CBD; ds.
XX
OS Clostridium cellulovorans.
XX
FH Key Location/Qualifiers
FT CDS 1..486
FT /*tag= a
XX
PN WO9424158-A.
XX
PD 27-OCT-1994.
XX
PF 14-APR-1994; 94WO-US04132.
XX
PR 14-APR-1993; 93US-0048164.
XX
PA (REGC ) UNIV CALIFORNIA.
PA (YISS ) YISSUM RES & DEV CO.
XX
PI Dol RH, Goldstein MA, Shoseyov O, Shplegl I;
XX
DR MPI: 1994-341767/42.
DR P-PSDB; AAR62634.
```

```
PT Isolated cellulose binding domain and fusion proteins - with
PT applications, including drug delivery, affinity separations, and
PT diagnostic techniques
XX
PS Claim 10; Fig 1; 125pp; English.
XX
CC AAQ72917/R63634 is a novel isolated cellulose binding domain. It pref.
CC binds crystalline with a Kd ranging from 1.5 microm - 0.5 microm,
CC pref. with a Kd of <1.2 microm, esp. <1.0 microm. The protein and
CC the encoding DNA are claimed, as is a CBD fusion protein (FP)
CC comprising the CBD and a second protein. The second protein is pref.
CC an environmental pollutant, or an HSP, HSP antibody, cross-reactive
CC HSP-related protein or an antigenic portion of this. The CBD and FP
CC may be used in drug delivery, affinity seps, and diagnostic
CC techniques. CBD nucleic acid may be obtd. from a variety of cell
CC sources that produce CBDs that bind with high affinity and in a
```

CC reversible manner or that produce CBD encoding mRNA. The preferred
 CC source of CBD encoding nucleic acid is C. cellulovorans.
 XX
 SQ Sequence 486 BP; 194 A; 89 C; 79 G; 124 T; 0 other;

Query Match 33.2%; Score 255; DB 15; Length 486;
 Best Local Similarity 100.0%; Pred. No. 7e-123;
 Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGACACAACAACCAACTCAATTACACCATTAATCAAAATTACTAAC 75
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 31 TACAACCTTAACAATGACACAACAACCAACTCAATTACACCATTAATCAAAATTACTAAC 90
 QY 76 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATTATTACACAAGTGAT 135
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 91 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATTATTACACAAGTGAT 150
 QY 136 GGTACACACAGACAACCTTCTGTGTGACCATGCTGTGCATTATTAGCAAAATAGCTAT 195
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 151 GGTACACACAGACAACCTTCTGTGTGACCATGCTGTGCATTATTAGCAAAATAGCTAT 210
 QY 196 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTGCTTAAAGAACGCAAGCCCAACATCA 255
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 211 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTGCTTAAAGAACGCAAGCCCAACATCA 270
 QY 256 ACCTATGATACATAT 270
 ||||||||||||||||||
 Db 271 ACCTATGATACATAT 285

RESULT 6
 AAV74072
 ID AAV74072 standard; DNA: 486 BP.
 XX
 AC AAV74072;
 XX
 DT 09-MAR-1999 (first entry)
 XX
 DE C. cellulovorans cbpa CBD DNA.
 XX
 KW Cellulose binding domain; CBD; cbpa; fusion protein; diagnosis; HSP;
 KW Immunoassay; heat-shock protein; cross reactive protein; detection;
 KW antigenic fragment; antibody; insulin-dependent diabetes mellitus;
 KW cellulose; chitin; cellulolytic; amorphogenic; ds.
 XX
 OS Clostridium cellulovorans.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..486
 FT /*tag= a
 FT /product= "cellulose binding domain"

XX US5856201-A.
 XX 05-JAN-1999.
 XX 27-OCT-1994; 94US-0330394.
 XX 27-OCT-1994; 94US-0330394.
 XX PR 14-APR-1993; 93US-0048164.
 XX PR 14-APR-1994; 94WO-US04132.
 XX (REGC) UNIV CALIFORNIA.
 XX (YISS) YISSUM RES & DEV CO.
 XX DoI RH, Goldstein MA, Shoseyov O, Shpiegl I;
 XX WPI: 1999-105130/09.
 XX P-PSDB: AAW90077.
 XX
 PT Detection of a specific analyte by reaction with binding agent fused
 PT to cellulose binding domain - and subsequent treatment with
 PT cellulose and reaction of insoluble product with a label specific

PT for the analyte
 XX
 XX Example 7.2.1; Fig 1A-B; 63pp; English.

CC This sequence encodes a cellulose binding domain (CBD) derived from the
 CC cbpa protein of Clostridium cellulovorans. The sequence is used the
 CC construction of a fusion protein which can be used in diagnostic
 CC immunoassays, e.g. to detect heat-shock proteins (HSP) and their
 CC cross-reactive proteins, antigenic fragments or HSP-specific antibodies
 CC (which indicate insulin-dependent diabetes mellitus, or susceptibility
 CC to it. The CBD, has very high affinity for cellulose (including
 CC crystalline forms) and chitin (dissociation constant 0.8-1.4 mu M), but
 CC has almost no cellulolytic or amorphogenic activities. The CBD binds over
 CC a wide pH range and is not released from cellulose by washing with water.
 XX
 SQ Sequence 486 BP; 194 A; 89 C; 79 G; 124 T; 0 other;

Query Match 33.2%; Score 255; DB 20; Length 486;
 Best Local Similarity 100.0%; Pred. No. 7e-123;
 Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGACACAACAACCAACTCAATTACACCATTAATCAAAATTACTAAC 75
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 31 TACAACCTTAACAATGACACAACAACCAACTCAATTACACCATTAATCAAAATTACTAAC 90
 QY 76 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATTATTACACAAGTGAT 135
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 91 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATTATTACACAAGTGAT 150
 QY 136 GGTACACACAGACAACCTTCTGTGTGACCATGCTGTGCATTATTAGCAAAATAGCTAT 195
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 151 GGTACACACAGACAACCTTCTGTGTGACCATGCTGTGCATTATTAGCAAAATAGCTAT 210
 QY 196 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTGCTTAAAGAACGCAAGCCCAACATCA 255
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 211 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTGCTTAAAGAACGCAAGCCCAACATCA 270
 QY 256 ACCTATGATACATAT 270
 ||||||||||||||||||
 Db 271 ACCTATGATACATAT 285

RESULT 7
 AAX24930
 ID AAX24930 standard; DNA: 499 BP.
 XX
 AC AAX24930;
 XX
 DT 21-JUN-1999 (first entry)
 XX
 DE Clostridium cellulovorans cbpa cellulose binding domain DNA.
 XX
 KW Cellulose binding domain; CBD; cbpa; endo-1,4-beta-glucanase; Cell;
 KW Arabidopsis thaliana; transgenic plant; crop improvement;
 KW morphology; cell wall; ds.
 XX
 OS Clostridium cellulovorans.
 XX
 PN WO9907830-A1.
 XX 18-FEB-1999.
 XX 26-JUL-1998; 98WO-1100345.
 XX 13-JAN-1998; 98US-0006636.
 XX PR 27-JUL-1997; 97IL-0121404.
 XX PR 13-JAN-1998; 98US-0006632.
 XX (YISS) YISSUM RES & DEV CO.
 XX Shani Z, Shoseyov O, Shpiegl E;
 XX WPI: 1999-180488/15.

XX Transgenic plants expressing cell-wall modulating protein - have
PT altered morphology, e.g. increased growth, modified fiber length or
PT cellulose content
PS Disclosure: Page 134; 144pp; English.
XX This DNA fragment encodes the cellulose binding domain (CBD) of
CC the cbpa protein of Clostridium cellulovorans. It was obtained
CC by PCR amplification (see also AX24952-53). and was used in the
CC construction of binary vector pCCT1 in which cbd was joined to
CC the promoter and signal region of the novel endo-1,4-beta-glucanase
CC cell gene (see AX24923) of Arabidopsis thaliana. Expression of cbd
CC in transgenic tobacco plants modulated their growth. The cell
CC promoter can provide expression of any protein in elongating
CC tissue. CBD is an example of a cell wall modulation transgene
CC used to alter the structure or morphology of a plant. Transgenic
CC plants of the invention may have altered biomass, growth, yield,
CC greater or less resistance to biodegradation, be more or less
CC digestible by ruminants, have altered cellulose content, larger or
CC smaller leaves, etc., when compared to non-transgenic plants of the
CC same species.
SQ Sequence 499 BP; 197 A; 93 C; 82 G; 127 T; 0 other;
Query Match 33.2%; Score 255; DB 20; Length 499;
Best Local Similarity 100.0%; Pred. No. 7e-123;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 TACACCTTACCAATTCAGCACAACCAATCTAATACCAATATCAAAATTACTTAC 75
DB 36 TACAACTCTTAACAAATTCAGCACAACCAATCTAATACCAATATCAAAATTACTTAC 95
QY 76 ACATCTGACAGTATTAAATTTAAATGACCTAAAGTTATATTATACCAAGTAT 135
DB 96 ACATCTGACAGTATTAAATTTAAATGACCTAAAGTTATATTATACCAAGTAT 155
QY 136 GGTGATTAACACTGACCAAACTTCTGTGACCAATGCTGTCATTATTAAGAAATAGCTAT 195
DB 156 GGTGATTAACACTGACCAAACTTCTGTGACCAATGCTGTCATTATTAAGAAATAGCTAT 215
QY 196 GTTGATTAACACTGACCAAACTTCTGTGACCAATGCTGTCATTATTAAGAAATAGCTAT 255
DB 216 GTTGATTAACACTGACCAAACTTCTGTGACCAATGCTGTCATTATTAAGAAATAGCTAT 275
QY 256 ACCTATGATACATAT 270
DB 276 ACCTATGATACATAT 290
RESULT 8
AAD11042 ID AAD11042 standard; DNA; 507 BP.
XX AAD11042:
XX 24-SEP-2001 (first entry)
XX Clostridium cellulovorans cellulose binding domain (CBDc1os) DNA.
DE
XX Polysaccharide modification; polysaccharide binding domain; PBD; paper;
KM yarn; fiber; textile; biological crosslinker; mechanical property;
KM Clostridium cellulovorans cellulose binding domain; wet strength;
KM durability; elasticity; CBDc1os; cellulose binding protein A; CBP A; ds.
XX
XX Clostridium cellulovorans.
OS
XX
FH Key Location/Qualifiers
FT CDS 3..494
FT /*tag= a
FT /product= "CBDc1os protein"
XX
PN MO200134091-A2.

XX 17-MAY-2001.
PD
XX
XX 02-NOV-2000; 2000WO-1100708.
PF
XX 08-NOV-1999; 99US-0164140.
PR 18-NOV-1999; 99US-0166389.
XX
PA (CBDT-) CBD TECHNOLOGIES LTD.
PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
PI Levy I, Nussinovitch A, Shoseyov O;
XX
XX WPI: 2001-457121/49.
DR P-PSDB; AAE05745.
XX
PT Preparation of a polysaccharide containing material having at least one
PT desired structural, chemical, physical, electrical and/or mechanical
PT property -
XX
PS Example 1.1; Page 111; 121pp; English.
XX
CC The present invention relates to methods and compositions for cross-
CC linking and/or modifying the properties of polysaccharide materials.
CC The method involves treating the polysaccharide structure with a
CC polysaccharide binding domain (PBD) fusion protein. The method is
CC used to alter the structural, chemical, physical, electrical and
CC mechanical properties of polysaccharide materials such as paper,
CC yarns, fibers and textiles, using biological crosslinking agents.
CC The polysaccharide containing materials have improved mechanical
CC properties such as wet strengths, durability and elasticity. The PBD
CC reagent is applied in the forming stage in fluting paper manufacture
CC which eliminates the sizing step. The use of a biological crosslinker
CC improves the recyclability of paper products. The PBD reagent maintains
CC the fine fibers in a slurry therefore resulting in better recovery of
CC raw materials. The PBD molecules are eluted by strong alkaline conditions
CC which enhances the ability of the alkaline glue used in binding to
CC penetrate paper.
CC The present sequence is a DNA encoding Clostridium cellulovorans
CC cellulose binding domain (CBDc1os) of cellulose binding protein A
CC (CBP A).
SQ Sequence 507 BP; 200 A; 94 C; 84 G; 129 T; 0 other;
Query Match 33.2%; Score 255; DB 22; Length 507;
Best Local Similarity 100.0%; Pred. No. 7e-123;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 TACAACTCTTAACAAATTCAGCACAACCAATCTAATACCAATATCAAAATTACTTAC 75
DB 36 TACAACTCTTAACAAATTCAGCACAACCAATCTAATACCAATATCAAAATTACTTAC 95
QY 76 ACATCTGACAGTATTAAATTTAAATGACCTAAAGTTATATTATACCAAGTAT 135
DB 96 ACATCTGACAGTATTAAATTTAAATGACCTAAAGTTATATTATACCAAGTAT 155
QY 136 GGTGATTAACACTGACCAAACTTCTGTGACCAATGCTGTCATTATTAAGAAATAGCTAT 195
DB 156 GGTGATTAACACTGACCAAACTTCTGTGACCAATGCTGTCATTATTAAGAAATAGCTAT 215
QY 196 GTTGATTAACACTGACCAAACTTCTGTGACCAATGCTGTCATTATTAAGAAATAGCTAT 255
DB 216 GTTGATTAACACTGACCAAACTTCTGTGACCAATGCTGTCATTATTAAGAAATAGCTAT 275
QY 256 ACCTATGATACATAT 270
DB 276 ACCTATGATACATAT 290
RESULT 9
AAD11046 ID AAD11046 standard; DNA; 984 BP.
XX

AC	ADD11046;	
XX		
DT	24-SEP-2001	(first entry)
XX		
DE	Chimeric S peptide-cellulose binding domain-S protein encoding DNA.	
XX		
KW	Polysaccharide modification; polysaccharide binding domain; PBD; paper;	
KW	ylan; fiber; textile; biological crosslinker; mechanical property;	
KW	wet strength; durability; elasticity; cellulose binding domain; CBD;	
KW	chimeric protein; S peptide-cellulose binding domain-S protein;	
KW	Spep-CBD-Sprot; bovine; ds.	
XX		
OS	Chimeric - Clostridium cellulovorans.	
OS	Chimeric - Bos sp.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	1..984
FT		/tag- a
FT		/product- "Spep-CBD-Sprot protein"
FT		/transl_except- "(pos: 979..981, aa:Xaa)"
FT		/note- "Xaa corresponds to in-frame stop codon;
FT		CDS does not include start and stop codon"
FT		/partial
FT	misc_feature	68..624
FT		//tag- b
FT		/note- "This region is derived from C. cellulovorans"
FT	misc_feature	652..981
FT		//tag- c
FT		/note- "This region is derived from bovine"
XX		
XX	W0200134091-A2.	
XX		
PD	17-MAY-2001.	
XX		
PE	02-NOV-2000;	2000MO-IL00708.
XX		
PR	08-NOV-1999;	9905-0164140.
PR	18-NOV-1999;	9905-0166389.
XX		
PA	(CBDT-) CBD TECHNOLOGIES LTD.	
PA	(YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.	
XX		
PI	Levy I, Nussinovitch A, Shoseyov O;	
XX		
DR	WPI: 2001-457121/49.	
DR	P-PSDB; AAE05749.	
XX		
PT	Preparation of a polysaccharide containing material having at least one	
PT	desired structural, chemical, physical, electrical and/or mechanical	
PT	property -	
XX		
PS	Example 4; Fig 4b-4g; 121pp; English.	
XX		
CC	The present invention relates to methods and compositions for cross-	
CC	linking and/or modifying the properties of polysaccharide materials.	
CC	The method involves treating the polysaccharide structure with a	
CC	polysaccharide binding domain (PBD) fusion protein. The method is	
CC	used to alter the structural, chemical, physical, electrical and	
CC	mechanical properties of polysaccharide materials such as paper,	
CC	yarns, fibers and textiles, using biological crosslinking agents.	
CC	The polysaccharide containing materials have improved mechanical	
CC	properties such as wet strengths, durability and elasticity. The PBD	
CC	reagent is applied in the forming stage in fluting paper manufacture	
CC	which eliminates the sizing step. The use of a biological crosslinker	
CC	improves the recyclability of paper products. The PBD reagent maintains	
CC	the fine fibers in a slurry therefore resulting in better recovery of	
CC	raw materials. The PBD molecules are eluted by strong alkaline conditions	
CC	which enhances the ability of the alkaline glue used in binding to	
CC	penetrate paper.	
CC	The present sequence is a DNA encoding S peptide-cellulose binding	
CC	domain-S protein (Spep-CBD-Sprot), a fusion protein derived from	
CC	Clostridium cellulovorans and bovine.	
XX		

Seq	Sequence	984 BP;	325 A;	240 C;	193 G;	226 T;	0 other;
QY	Query Match	33.2%;	Score 255;	DB 22;	Length 984;		
Db	Best Local Similarity	100.0%;	Pred. No. 7e-123;				
Matches	255;	Conservative	0;	Mismatches	0;	Indels	0; Gaps
QY	16	TCAACTCTAACAAATCAGCACACAAACAAACTCAATTACACCAATTAATCAAAATTACTAAC	75				
Db	118	TACAACCTCTAACAAATCAGCACACAAACAAACTCAATTACACCAATTAATCAAAATTACTAAC	177				
QY	76	ACATCTGCAGAGGATTTAAATTTAAATGACGTAAAGTTAGATATTATTATACACAAGGAT	135				
Db	178	ACATCTGCAGAGGATTTAAATTTAAATGACGTAAAGTTAGATATTATTATACACAAGGAT	237				
QY	136	GGTACACAAAGGACAAACTTCTGCTGTGACACATGCTGTGCATATTATTAGAAATAGCTAT	195				
Db	238	GGTACACAAAGGACAAACTTCTGCTGTGACACATGCTGTGCATATTATTAGAAATAGCTAT	297				
QY	196	GTTGATACACCTAGCACAAGTGACAGCAAACTTCGTTAAAGAAACAGCAGCCACATCA	255				
Db	298	GTTGATACACCTAGCACAAGTGACAGCAAACTTCGTTAAAGAAACAGCAGCCACATCA	357				
QY	256	ACCTATGATACATAT 270					
Db	358	ACCTATGATACATAT 372					
RESULT 10							
AAZ27550							
ID	AAZ27550	standard;	DNA;	1146	BP.		
XX	AAZ27550;						
AC							
XX							
DT	13-DEC-1999	(first entry)					
XX							
XX	Gaussia luciferase fusion protein coding sequence.						
DE							
KM	Luciferase; green fluorescent protein; GFP; screening assay; diagnosis;						
KM	bioluminescence-generating system; toy; cosmetic; fairy dust; beverage;						
KM	body paint; squirt gun; balloon; slimy play material; soap; toothpaste;						
KM	fusion protein; ds.						
XX							
OS	Gaussia sp.						
XX							
PN	W09949019-A2.						
XX							
PD	30-SEP-1999.						
XX							
PF	26-MAR-1999;	99WO-US06698.					
XX							
PR	27-MAR-1998;	98US-0079624.					
PR	15-JUN-1998;	98US-0089367.					
PR	01-OCT-1998;	98US-0102939.					
XX							
PA	(PROL-) PROLUME LTD.						
XX	(BRYA-) BRYAN B J.						
XX							
PI	Bryan BJ, Szent-Gyorgyi C;						
XX							
XX							
XX	WPI; 1999-580443/49.						
DR	P-PSDB; AAY39952.						
PT							
XX	New isolated Renilla mulleri, Gaussia and Pleuromamma luciferase and						
XX	Renilla and Ptilosarcus green fluorescent protein nucleic acids -						
PS	Disclosure; Page 222-223; 233pp; English.						
XX							
XX	This sequence encodes a luciferase of the invention. The invention						
CC	relates to Renilla mulleri, Gaussia and Pleuromamma luciferase and						
CC	Renilla and Ptilosarcus green fluorescent protein (GFP) nucleic acids and						
CC	proteins. The luciferases and GFPs can be used in						
CC	bioluminescence-generating systems, assays, screening methods, diagnostic						
CC	method and articles of manufacture. They can be expressed using						

```
CC e.g. bacterial, yeast, fungal, plant, insect or animal cells. The Renilla
CC milleri, Gaussia and Pleuromamma luciferase or Renilla or Ptilosarcus
CC GFPs can be used in e.g. toys, cosmetics, fountain, personal care items,
CC fairy dust, beverages, soft drinks, foods, textile products, bubbles,
CC balloons, personal items, dentifrices, soaps, body paints, bubble bath,
CC ink or paper products. In particular, they can be used in e.g. squirt
CC guns, pellet guns, finger paints, foot bags, greeting cards, slimy play
CC material, clothing, bubble making toys, bath powders, cosmetics, body
CC lotions, gels, body powders, body creams, toothpastes, mouthwashes,
CC soaps, body paints, bubble bath, inks, wrapping paper, gelatins, icings,
CC frostings, greeting cards, beer, wine, champagne, soft drinks, ice cubes,
CC ice, dry ice or fountain. The nucleic acids can also be used to produce
CC transgenic fish and plants.
XX
XX
SQ Sequence 1146 BP; 396 A; 222 C; 247 G; 281 T; 0 other;

Query Match          33.2%; Score 255; DB 20; Length 1146;
Best Local Similarity 100.0%; Pred. No. 7e-123;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 TACAACCTCTAACAAATGACGACAAACAACTCAATTACACCAATATCAAAATTACTAAC 75
    |||||||
DB 16 TACAACCTCTAACAAATGACGACAAACAACTCAATTACACCAATATCAAAATTACTAAC 75
OY 76 ACATCTGACAGTGATTAATTTAAATGACGTAAGTATTTATACACAGATGAT 135
    |||||||
DB 76 ACATCTGACAGTGATTAATTTAAATGACGTAAGTATTTATACACAGATGAT 135
OY 136 GGTACACAGAGCAAACTTTCTGCTGTGACCATCTGCTGTCATTATTAGCAATAGCTAT 195
    |||||||
DB 136 GGTACACAGAGCAAACTTTCTGCTGTGACCATCTGCTGTCATTATTAGCAATAGCTAT 195
OY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAAGCAAGCCCAACATCA 255
    |||||||
DB 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAAGCAAGCCCAACATCA 255
OY 256 ACCTATGATACATAT 270
    |||||||
DB 256 ACCTATGATACATAT 270

RESULT 11
AAD22201 standard; DNA; 1146 BP.
XX
AC AAD22201;
XX
DT 12-FEB-2002 (first entry)
XX
DE Gaussia species CBD-luciferase fusion protein encoding DNA.
XX
KW Green fluorescent protein; GFP; bioluminescence generating system; toy;
KW luciferase; finger paint; slimy play material; fishing lure; sparkler;
KW doll; balloon; personal care item; cosmetic; bath powder; body cream;
KW tooth paste; mouth wash; soap; body paint; ornamental transgenic plant;
KW bubble bath; fountain; fairy dust; food; gelatin; icing; frosting; beer;
KW wine; champagne; milk; soft drink; ice cube; ice; foot bag; beverage;
KW dry ice; textile; clothing; paper product; greeting card; wrapping paper;
KW fusion protein; ds.
XX
OS Gaussia sp.
XX
XX
FH Key 1. 1146 Location/Qualifiers
FT CDS
FT /tag= a
FT /product= "Gaussia CBD-luciferase fusion protein"
FT /note= "CDS does not include start and stop codon"
FT /partial
XX
XX WO200168824-A2.
XX
PD 20-SEP-2001.
XX
```

```
PF 15-MAR-2001; 2001WO-US08227.
XX
XX 15-MAR-2000; 2000US-189691P.
XX
XX (PROL-) PROLUME LTD.
XX (BRYA-) BRYAN B J.
XX
XX Bryan BJ, Szent-Gyorgyi C, Szczepaniak W;
XX WPI, 2002-010561/01;
XX P-PSDB: AAE1383.
XX
XX Nucleic acids encoding Renilla reniformis green fluorescent proteins,
XX useful in diagnostic bioluminescence procedures -
XX
XX Disclosure; Page 162-163; 175pp; English.
XX
XX The patent discloses sea pansy (Renilla reniformis) green fluorescent
XX proteins (GFP) and their corresponding polynucleotides. The invention
XX also relates to sequences of the bioluminescence generating system
XX (e.g. luciferase). R. reniformis GFP are used in diagnostic methods
XX and in the production of novelty items such as toys (e.g. squirt gun,
XX pellet gun, toy "Halloween" eggs, toy cigarettes, board/card game
XX toys), finger paints, slimy play material, bubbles in bubble making
XX CC toys, fishing lures, dolls, sparklers, magic wand toys, balloons,
XX CC personal care item (e.g. cosmetic, bath powders, body creams, tooth
XX CC pastes, mouth wash, soaps, body paints, bubble bath), ornamental
XX CC transgenic plants, fountain, fairy dust, food (gelatins, icings,
XX CC frosting, beers, wine, champagne, milk, soft drinks, ice cubes, ice,
XX CC dry ice, beverage), textile (foot bag, clothing) and/or paper product
XX CC (greeting cards, wrapping paper). The present sequence is a DNA
XX encoding Gaussia species CBD-luciferase fusion protein.
XX
XX
SQ Sequence 1146 BP; 396 A; 222 C; 247 G; 281 T; 0 other;

Query Match          33.2%; Score 255; DB 24; Length 1146;
Best Local Similarity 100.0%; Pred. No. 7e-123;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 TACAACCTCTAACAAATGACGACAAACAACTCAATTACACCAATATCAAAATTACTAAC 75
    |||||||
DB 16 TACAACCTCTAACAAATGACGACAAACAACTCAATTACACCAATATCAAAATTACTAAC 75
OY 76 ACATCTGACAGTGATTAATTTAAATGACGTAAGTATTTATACACAGATGAT 135
    |||||||
DB 76 ACATCTGACAGTGATTAATTTAAATGACGTAAGTATTTATACACAGATGAT 135
OY 136 GGTACACAGAGCAAACTTTCTGCTGTGACCATCTGCTGTCATTATTAGCAATAGCTAT 195
    |||||||
DB 136 GGTACACAGAGCAAACTTTCTGCTGTGACCATCTGCTGTCATTATTAGCAATAGCTAT 195
OY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAAGCAAGCCCAACATCA 255
    |||||||
DB 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAAGCAAGCCCAACATCA 255
OY 256 ACCTATGATACATAT 270
    |||||||
DB 256 ACCTATGATACATAT 270

RESULT 12
AAD11045 standard; DNA; 1288 BP.
XX
AC AAD11045;
XX
DT 24-SEP-2001 (first entry)
XX
DE Clostridium cellulovorans protein A-cellulose binding domain DNA.
XX
XX Polysaccharide modification; polysaccharide binding domain; PBD; paper;
KW yarn; fiber; textile; biological crosslinker; mechanical property;
KW wet strength; durability; elasticity; protein A-cellulose binding domain;
```

```

KM ProA-CBD; ds.
XX
XX Clostridium cellulovorans.
XX
FH Key Location/Qualifiers
FT 3..1286
FT /tag= a
FT /product= "Protein A-cellulose binding domain protein"
FT /transl_except= "(pos: 1281..1283, aa: xaa)"
FT /note= "Xaa corresponds to in-frame stop codon:
FT CDS does not include stop codon"
FT /partial
FT misc_feature
FT 3..791
FT /tag= b
FT /note= "This region is derived from PRIT2T
FT Cloning vector"
FT misc_feature
FT 795..1280
FT /tag= c
FT /note= "This region is derived from cbp A gene"
XX
XX WO200134091-A2.
XX
XX 17-MAY-2001.
XX
XX 02-NOV-2000; 2000MO-IL00708.
XX
XX 08-NOV-1999; 9905-0164140.
XX 18-NOV-1999; 9905-0166389.
XX
XX (CBDT-) CBD TECHNOLOGIES LTD.
XX (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
XX Levy I, Nussinovitch A, Shoseyov O;
XX
XX WPI; 2001-457121/49.
XX P-PSDB; AAE05748.
XX
XX Preparation of a polysaccharide containing material having at least one
XX desired structural, chemical, physical, electrical and/or mechanical
XX property.
XX
XX Example 1.3; Fig 3b-3g; 121pp; English.
XX
XX The present invention relates to methods and compositions for cross-
XX linking and/or modifying the properties of polysaccharide materials.
XX The method involves treating the polysaccharide structure with a
XX polysaccharide binding domain (PBD) fusion protein. The method is
XX used to alter the structural, chemical, physical, electrical and
XX mechanical properties of polysaccharide materials such as paper,
XX yarns, fibers and textiles, using biological crosslinking agents.
XX The polysaccharide containing materials have improved mechanical
XX properties such as wet strengths, durability and elasticity. The PBD
XX reagent is applied in the forming stage in fluting paper manufacture
XX which eliminates the sizing step. The use of a biological crosslinker
XX improves the recyclability of paper products. The PBD reagent maintains
XX the fine fibers in a slurry therefore resulting in better recovery of
XX raw materials. The PBD molecules are eluted by strong alkaline conditions
XX which enhances the ability of the alkaline glue used in binding to
XX penetrate paper.
XX The present sequence is a DNA encoding Clostridium cellulovorans
XX Protein A-cellulose binding domain (ProA-CBD). This sequence contains
XX a part of cbpA gene and a part of PRIT2T vector sequence.
XX
XX Sequence 1288 BP; 523 A; 267 C; 202 G; 296 T; 0 other;
XX
XX Query Match 33.2%; Score 255; DB 22; Length 1288;
XX Best Local Similarity 100.0%; Pred. No. 7e-123;
XX Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 76 ACATCTGACAGTGTAAATTAATGACGTAAAGATAGTATTATTACACAGAT 135
XX |
XX |
XX |
DB 885 ACATCTGACAGTGTAAATTAATGACGTAAAGATAGTATTATTACACAGAT 944
OY 136 GGTACACAGACAACTTCTGTGTGACCATGCTGTCATTATTTAGAAATAGCTAT 195
XX |
XX |
XX |
DB 945 GGTACACAGACAACTTCTGTGTGACCATGCTGTCATTATTTAGAAATAGCTAT 1004
OY 196 GTTGATATACACTGTGCAAGTGACGCAAACTTGTTAAGAAACAGCAGCCCAACATCA 255
XX |
XX |
XX |
DB 1005 GTTGATATACACTGTGCAAGTGACGCAAACTTGTTAAGAAACAGCAGCCCAACATCA 1064
OY 256 ACCTATGATACATAT 270
XX |
XX |
DB 1065 ACCTATGATACATAT 1079
XX
XX RESULT 13
XX AAF86254/c
XX ID AAF86254 standard; DNA; 118 BP.
XX
XX AAF86254;
XX
XX 11-JUL-2001 (first entry)
XX
XX PCR primer #4 used in cloning an optimisation of OspA gene.
XX
XX Polkilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
XX KW vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
XX SRS; 17E2; fusion construct; PCR primer; ss.
XX
XX Piscirickettsia salmonis.
XX
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
XX (BURJ/) BURIAN J.
XX (KUZV/) KUZYZK M A.
XX
XX Kay WW, Burian J, Kuzyzk MA;
XX
XX WPI; 2001-316844/34.
XX
XX Method for protecting polkilothermic fish against salmonid rickettsial
XX septicaemia and other rickettsial diseases comprises administering a
XX vaccine containing the OspA protein of Piscirickettsia salmonis.
XX
XX Example 3; Fig 4B; 35pp; English.
XX
XX This invention relates to a method for the protection against infection
XX of a polkilothermic fish by the bacterial pathogen, Piscirickettsia
XX salmonis. The method comprises administering an immunogenic amount of a
XX P. salmonis specific antigen termed OspA, or an immunogenic fragment of
XX OspA in the form of a vaccine. The method is used for protecting animals,
XX particularly polkilothermic fish, against the bacterial pathogen
XX P. salmonis. The method is also useful for protecting against salmonid
XX rickettsial septicaemia (SRS) and other rickettsial diseases. The present
XX sequence represents a PCR primer used in the cloning and optimisation of
XX the P. salmonis OspA gene. The OspA gene is used in the method of the
XX invention.
XX
XX Sequence 118 BP; 21 A; 32 C; 28 G; 37 T; 0 other;
XX
XX Query Match 15.4%; Score 118; DB 22; Length 118;
XX Best Local Similarity 100.0%; Pred. No. 2.4e-51;
XX Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 493 GGTGAGATGACGACGAGATTAATCAACGACGCTCTCTGGAAGTGAAG 552
|
DB 118 GGTGAGATGACGACGAGATTAATCAACGACGCTCTCTGGAAGTGAAG 59
QY 553 GCGGCGAGGTACTGTTGCGGATCCGACACCGGTAACTACTCTGTGGAAC 610
|
DB 58 GCGGCGAGGTACTGTTGCGGATCCGACACCGGTAACTACTCTGTGGAAC 1

RESULT 14

AAAF86252
ID AAAF86252 standard; DNA: 110 BP.

AC AAAF86252;

DT 11-JUL-2001 (first entry)

DE PCR primer #2 used in cloning an optimisation of OspA gene.

XX polkilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;

KW vaccine; OspA; salmonid rickettsial septicemia; rickettsial disease;

XX SRS; 17E2; fusion construct; PCR primer; ss.

OS Piscirickettsia salmonis.

PN CA2281913-A1.

PD 17-MAR-2001.

PE 17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

PA (KAYW/) KAY W W.

PA (BURI/) BURIAN J.

PA (KUZY/) KUZUK M A.

PI Kay WW, Burian J, Kuzuk MA;

DR WPI: 2001-316844/34.

XX Method for protecting polkilothermic fish against salmonid rickettsial

PT septicemia and other rickettsial diseases comprises administering a

PT vaccine containing the OspA protein of Piscirickettsia salmonis

XX Example 3; Fig 4B; 35pp; English.

XX This invention relates to a method for the protection against infection

CC of a polkilothermic fish by the bacterial pathogen, Piscirickettsia

CC P. salmonis. The method comprises administering an immunogenic fragment of a

CC P. salmonis specific antigen termed OspA, or an immunogenic fragment of

CC OspA in the form of a vaccine. The method is used for protecting animals,

CC particularly polkilothermic fish, against the bacterial pathogen

CC P. salmonis. The method is also useful for protecting against salmonid

CC rickettsial septicemia (SRS) and other rickettsial diseases. The present

CC sequence represents a PCR primer used in the cloning and optimisation of

CC the P. salmonis OspA gene. The OspA gene is used in the method of the

CC invention.

XX Sequence 110 BP; 13 A; 28 C; 45 G; 24 T; 0 other;

QY Query Match 13.5%; Score 104; DB 22; Length 110;

DB Best Local Similarity 100.0%; Pred. No. 4.8e-44;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 15
AAAF86252/C
ID AAAF86252 standard; DNA: 102 BP.

AC AAAF86252;

DT 11-JUL-2001 (first entry)

DE PCR primer #5 used in cloning an optimisation of OspA gene.

XX polkilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;

KW vaccine; OspA; salmonid rickettsial septicemia; rickettsial disease;

XX SRS; 17E2; fusion construct; PCR primer; ss.

OS Piscirickettsia salmonis.

PN CA2281913-A1.

PD 17-MAR-2001.

PE 17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

PA (KAYW/) KAY W W.

PA (BURI/) BURIAN J.

PA (KUZY/) KUZUK M A.

PI Kay WW, Burian J, Kuzuk MA;

DR WPI: 2001-316844/34.

XX Method for protecting polkilothermic fish against salmonid rickettsial

PT septicemia and other rickettsial diseases comprises administering a

PT vaccine containing the OspA protein of Piscirickettsia salmonis

XX Example 3; Fig 4B; 35pp; English.

XX This invention relates to a method for the protection against infection

CC of a polkilothermic fish by the bacterial pathogen, Piscirickettsia

CC P. salmonis. The method comprises administering an immunogenic amount of a

CC P. salmonis specific antigen termed OspA, or an immunogenic fragment of

CC OspA in the form of a vaccine. The method is used for protecting animals,

CC particularly polkilothermic fish, against the bacterial pathogen

CC P. salmonis. The method is also useful for protecting against salmonid

CC rickettsial septicemia (SRS) and other rickettsial diseases. The present

CC sequence represents a PCR primer used in the cloning and optimisation of

CC the P. salmonis OspA gene. The OspA gene is used in the method of the

CC invention.

XX Sequence 102 BP; 14 A; 23 C; 30 G; 35 T; 0 other;

QY Query Match 13.3%; Score 102; DB 22; Length 102;

DB Best Local Similarity 100.0%; Pred. No. 5.4e-43;

Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 619 ACCTACGAGCGTTACAAACAGAGAGCGGTCAGCAGTACTGCGGATTTACGACG 678

DB 102 ACCTACGAGCGTTACAAACAGAGAGCGGTCAGCAGTACTGCGGATTTACGACG 43

QY 679 AAGGCATGATGCGAGGTCAGAAACAGAGAAATCTACGACGACC 720

DB 42 AAGGCATGATGCGAGGTCAGAAACAGAGAAATCTACGACGACC 1

Search completed: October 27, 2002, 20:42:33

Job time : 197.542 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 18:35:09 : Search time 1486.04 Seconds
(without alignments)
10815.053 Million cell updates/sec

Title: US-09-677-374-5

Perfect score: 768
Sequence: 1 atgtcagttggaattctacaa.....agggtgattagaccgaaaaa 768

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
------------	-------------	-------	--------	-------	-------------

1	255	33.2	486	6	AR001082	AR001082 Sequence
2	255	33.2	486	6	AR001083	AR001083 Sequence
3	255	33.2	486	6	AR027448	AR027448 Sequence
4	255	33.2	486	6	AR027449	AR027449 Sequence
5	255	33.2	486	6	AR058257	AR058257 Sequence
6	255	33.2	486	6	AR058258	AR058258 Sequence
7	255	33.2	486	6	118514	118514 Sequence 1
8	255	33.2	486	6	118515	118515 Sequence 3
9	255	33.2	486	6	166614	166614 Sequence 1
10	255	33.2	486	6	166615	166615 Sequence 3
11	255	33.2	486	6	188789	188789 Sequence 1
12	255	33.2	486	6	188790	188790 Sequence 3
13	255	33.2	486	6	AR096204	AR096204 Sequence
14	255	33.2	1146	6	AR051733	AR051733 Sequence
15	255	33.2	1146	6	AX250579	AX250579 Sequence
16	255	33.2	6760	1	CLOCBP	M73817 C10stridium
17	23	3.0	243340	2	AL663088	AL663088 MUS muscu
18	22	2.9	589	8	HJRNA1956	AJ001956 HYlomecon
19	22	2.9	594	8	CMRNA1959	AJ001959 Chelidoni
20	22	2.9	596	8	CMRNA1958	AJ001958 Chelidoni
21	22	2.9	596	8	CMRNA1960	AJ001960 Chelidoni
22	22	2.9	648	8	AF169233	AF169233 Zizania t
23	22	2.9	7286	1	AE005332	AE005332 Escherich
24	22	2.9	10029	1	AE005325	AE005325 Escherich
25	22	2.9	11809	1	AE005290	AE005290 Escherich
26	22	2.9	13505	1	AE005441	AE005441 Escherich
27	22	2.9	13624	1	AE006459	AE006459 Escherich
28	22	2.9	103608	9	AL139161	AL139161 Human DNA
29	22	2.9	222605	1	AP002555	AP002555 Escherich
30	22	2.9	286485	1	AP002557	AP002557 Escherich
31	22	2.9	291136	1	AP002556	AP002556 Escherich
32	21	2.7	702	8	AGHGBEHA	L13954 Agoseris he
33	21	2.7	705	8	MITGBEHA	L13955 Microseris
34	21	2.7	706	8	KIRGBEHA	L20464 Krigia bifl
35	21	2.7	706	8	KIRGBEHA	L13948 Krigia bifl
36	21	2.7	706	8	LAURGBEHA	L13957 Lactuca sat
37	21	2.7	708	8	AF422126	AF422126 Microseri
38	21	2.7	748	8	MCU69708	U69708 Microseris
39	21	2.7	750	8	AF386496	AF386496 Microseri
40	21	2.7	750	8	AF386497	AF386497 Microseri
41	21	2.7	750	8	AF386498	AF386498 Microseri
42	21	2.7	750	8	AF386499	AF386499 Microseri
43	21	2.7	750	8	AF386500	AF386500 Microseri
44	21	2.7	750	8	AF386501	AF386501 Microseri
45	21	2.7	750	8	AF386502	AF386502 Microseri

ALIGNMENTS

RESULT 1
AR001082
LOCUS AR001082 486 bp DNA
DEFINITION Sequence 1 from patent US 5738984.
ACCESSION AR001082
VERSION AR001082.1 GI:3963149
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov O.
TITLE Kits and methods of detection using cellulose binding domain fusion
JOURNAL Patent: US 5738984-A 1 14-APR-1998;
FEATURES Location/Qualifiers
source 1..486
/organism="unknown"

BASE COUNT	194 a	89 c	79 g	124 t
Query Match	33.2%	Score 255;	DB 6;	Length 486;
Best Local Similarity	100.0%;	Pred. No. 2.3e-139;		

Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATCAGCACAACAACTCAATTACACCAATTAATCAAAATTAATAAC 75
|||||
Db 31 TACAACCTTAACAATCAGCACAACAACTCAATTACACCAATTAATCAAAATTAATAAC 90

QY 76 ACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAAGTTGATTTATTTACACAAGTAT 135
|||||
Db 91 ACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAAGTTGATTTATTTACACAAGTAT 150

QY 136 GGTACACAAGGACAACCTTCTGTTGACCAAGCTGTCGATTTATTTAGGAATAGCTAT 195
|||||
Db 151 GGTACACAAGGACAACCTTCTGTTGACCAAGCTGTCGATTTATTTAGGAATAGCTAT 210

QY 196 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTCTGTTAAGAAACGACAGCCCAACATCA 255
|||||
Db 211 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTCTGTTAAGAAACGACAGCCCAACATCA 270

QY 256 ACCTATGATACATAT 270
|||||
Db 271 ACCTATGATACATAT 285

RESULT 2
AR001083/c AR001083 486 bp DNA linear PAT 04-DEC-1998
LOCUS Sequence 3 from patent US 5738984.
DEFINITION AR001083
ACCESSION AR001083
VERSION AR001083.1 GI:3963150
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 486)
Shoseyov, O.
TITLE Kits and methods of detection using cellulose binding domain fusion proteins
JOURNAL Patent: US 5738984-A 3 14-APR-1998;
FEATURES
source Location/Qualifiers
BASE COUNT 124 a 79 c 89 g 194 t
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;
Best Local Similarity 100.0%; Pred. No. 2.3e-139; Indels 0; Gaps 0;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATCAGCACAACAACTCAATTACCAATTAATCAAAATTAATAAC 75
|||||
Db 456 TACAACCTTAACAATCAGCACAACAACTCAATTACCAATTAATCAAAATTAATAAC 397

QY 76 ACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAAGTTGATTTATTTACACAAGTAT 135
|||||
Db 396 ACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAAGTTGATTTATTTACACAAGTAT 337

QY 136 GGTACACAAGGACAACCTTCTGTTGACCAAGCTGTCGATTTATTTAGGAATAGCTAT 195
|||||
Db 336 GGTACACAAGGACAACCTTCTGTTGACCAAGCTGTCGATTTATTTAGGAATAGCTAT 277

QY 196 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTCTGTTAAGAAACGACAGCCCAACATCA 255
|||||
Db 276 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTCTGTTAAGAAACGACAGCCCAACATCA 217

QY 256 ACCTATGATACATAT 270
|||||
Db 216 ACCTATGATACATAT 202

RESULT 3
AR027448 AR027448 486 bp DNA linear PAT 29-SEP-1999
LOCUS Sequence 1 from patent US 5856201.
DEFINITION

ACCESSION AR027448
VERSION AR027448.1 GI:5938268
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 486)
Shoseyov, O., Shpiegel, I., Goldstein, M.A. and Dol, R.H.
TITLE Methods of detection using a cellulose binding domain fusion product
JOURNAL Patent: US 5856201-A 1 05-JAN-1999;
FEATURES
source Location/Qualifiers
BASE COUNT 194 a 89 c 79 g 124 t
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;
Best Local Similarity 100.0%; Pred. No. 2.3e-139; Indels 0; Gaps 0;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATCAGCACAACAACTCAATTACCAATTAATCAAAATTAATAAC 75
|||||
Db 31 TACAACCTTAACAATCAGCACAACAACTCAATTACCAATTAATCAAAATTAATAAC 90

QY 76 ACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAAGTTGATTTATTTACACAAGTAT 135
|||||
Db 91 ACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAAGTTGATTTATTTACACAAGTAT 150

QY 136 GGTACACAAGGACAACCTTCTGTTGACCAAGCTGTCGATTTATTTAGGAATAGCTAT 195
|||||
Db 151 GGTACACAAGGACAACCTTCTGTTGACCAAGCTGTCGATTTATTTAGGAATAGCTAT 210

QY 196 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTCTGTTAAGAAACGACAGCCCAACATCA 255
|||||
Db 211 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTCTGTTAAGAAACGACAGCCCAACATCA 270

QY 256 ACCTATGATACATAT 270
|||||
Db 271 ACCTATGATACATAT 285

RESULT 4
AR027449/c AR027449 486 bp DNA linear PAT 29-SEP-1999
LOCUS Sequence 3 from patent US 5856201.
DEFINITION AR027449
ACCESSION AR027449
VERSION AR027449.1 GI:5938269
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 486)
Shoseyov, O., Shpiegel, I., Goldstein, M.A. and Dol, R.H.
TITLE Methods of detection using a cellulose binding domain fusion product
JOURNAL Patent: US 5856201-A 3 05-JAN-1999;
FEATURES
source Location/Qualifiers
BASE COUNT 124 a 79 c 89 g 194 t
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;
Best Local Similarity 100.0%; Pred. No. 2.3e-139; Indels 0; Gaps 0;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATCAGCACAACAACTCAATTACCAATTAATCAAAATTAATAAC 75
|||||
Db 456 TACAACCTTAACAATCAGCACAACAACTCAATTACCAATTAATCAAAATTAATAAC 397

QY 76 ACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAAGTTGATTTATTTACACAAGTAT 135
|||||

Db 396 ACATCTGACAGTATTTAAATTTAAATGACGTAAAGCTAGATATATATACACAGTGTAT 337
QY 136 GGTACACAAAGACAAACTTCTGCTGACCACTGCTGCATATATAGAAATAGCTAT 195
Db 336 GGTACACAAAGACAAACTTCTGCTGACCACTGCTGCATATATAGAAATAGCTAT 277
QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 255
Db 276 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 217
QY 256 ACCTATGATACATAT 270
Db 216 ACCTATGATACATAT 202

RESULT 5
AR058257
LOCUS AR058257 486 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5837814.
ACCESSION AR058257
VERSION AR058257.1 GI:5983834
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov,O., Shpiegl,I., Goldstein,M. and Dol,R.
TITLE Cellulose binding domain proteins
JOURNAL Patent: US 5837814-A 1 17-NOV-1998;
FEATURES
source 1..486
BASE COUNT 194 a 89 c 79 g 124 t
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;
Best Local Similarity 100.0%; Pred. No. 2.3e-139;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGACGACAAACAACTCAATTAACCAATATCAAAATTAAGTAC 75
Db 31 TACAACCTTAACAATGACGACAAACAACTCAATTAACCAATATCAAAATTAAGTAC 90
QY 76 ACATCTGACAGTATTTAAATTTAAATGAGCTAAAGTATTTATTTACACAGTGTAT 135
Db 91 ACATCTGACAGTATTTAAATTTAAATGAGCTAAAGTATTTATTTACACAGTGTAT 150
QY 136 GGTACACAAAGACAAACTTCTGCTGACCACTGCTGCATATATAGAAATAGCTAT 195
Db 151 GGTACACAAAGACAAACTTCTGCTGACCACTGCTGCATATATAGAAATAGCTAT 210
QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 255
Db 211 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 270
QY 256 ACCTATGATACATAT 270
Db 271 ACCTATGATACATAT 285

RESULT 6
AR058258/c
LOCUS AR058258 486 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 3 from patent US 5837814.
ACCESSION AR058258
VERSION AR058258.1 GI:5983835
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov,O., Shpiegl,I., Goldstein,M. and Dol,R.
TITLE Cellulose binding domain proteins

JOURNAL Patent: US 5837814-A 3 17-NOV-1998;
FEATURES
source 1..486
BASE COUNT 124 a 79 c 89 g 194 t
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;
Best Local Similarity 100.0%; Pred. No. 2.3e-139;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGACGACAAACAACTCAATTAACCAATATCAAAATTAAGTAC 75
Db 456 TACAACCTTAACAATGACGACAAACAACTCAATTAACCAATATCAAAATTAAGTAC 397
QY 76 ACATCTGACAGTATTTAAATTTAAATGAGCTAAAGTATTTATTTACACAGTGTAT 135
Db 396 ACATCTGACAGTATTTAAATTTAAATGAGCTAAAGTATTTATTTACACAGTGTAT 337
QY 136 GGTACACAAAGACAAACTTCTGCTGACCACTGCTGCATATATAGAAATAGCTAT 195
Db 336 GGTACACAAAGACAAACTTCTGCTGACCACTGCTGCATATATAGAAATAGCTAT 277
QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 255
Db 276 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 217
QY 256 ACCTATGATACATAT 270
Db 216 ACCTATGATACATAT 202

RESULT 7
I18514
LOCUS I18514 486 bp DNA linear PAT 07-OCT-1996
DEFINITION Sequence 1 from patent US 5496934.
ACCESSION I18514
VERSION I18514.1 GI:1598869
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Dol,R.H.
TITLE Nucleic acids encoding a cellulose binding domain
JOURNAL Patent: US 5496934-A 1 05-MAR-1996;
FEATURES
source 1..486
BASE COUNT 194 a 89 c 79 g 124 t
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;
Best Local Similarity 100.0%; Pred. No. 2.3e-139;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGACGACAAACAACTCAATTAACCAATATCAAAATTAAGTAC 75
Db 31 TACAACCTTAACAATGACGACAAACAACTCAATTAACCAATATCAAAATTAAGTAC 90
QY 76 ACATCTGACAGTATTTAAATTTAAATGAGCTAAAGTATTTATTTACACAGTGTAT 135
Db 91 ACATCTGACAGTATTTAAATTTAAATGAGCTAAAGTATTTATTTACACAGTGTAT 150
QY 136 GGTACACAAAGACAAACTTCTGCTGACCACTGCTGCATATATAGAAATAGCTAT 195
Db 151 GGTACACAAAGACAAACTTCTGCTGACCACTGCTGCATATATAGAAATAGCTAT 210
QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 255
Db 211 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 270
QY 256 ACCTATGATACATAT 270

Db 271 |||||
 ACCTATGATAACATAT 285

RESULT 8			
118515/c			
LOCUS	118515	486 bp	DNA
DEFINITION	Sequence 3 from patent US 5496934.		Linear
			PAT 07-OCT-1996

	/organism="unknown"			
BASE COUNT	124 a	79 c	89 g	194 t
ORIGIN				

Query Match	33.2%	Score 255;	DB 6;	Length 486;
Best Local Similarity	100.0%	Pred. No. 2.3e-139;		
Matches 255;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	16	TACAACTGTACAAATGCGACAAACAACTCAATTACACCAATTAACAAATACAC	75
Db	456	TACAACTGTACAAATGCGACAAACAACTCAATTACACCAATTAACAAATACAC	397
Qy	76	ACATCTGCAGTGAATTTAAATTTAAATGCGTAAAGTTAGATATTTACACAACTGAT	135
Db	396	ACATCTGCAGTGAATTTAAATTTAAATGCGTAAAGTTAGATATTTACACAACTGAT	337
Qy	136	GSTACACAAAGACAAACTTCTCGTGAGCAACATCGTGGCATATTATAGAAATAGCTAT	195
Db	336	GSTACACAAAGACAAACTTCTCGTGAGCAACATCGTGGCATATTATAGAAATAGCTAT	277
Qy	196	GTTGATTAACACTAGCAAAAGTACAGCAAACTTCGTTAAAGAAACGACGAAGCCCAACATCA	255
Db	276	GTTGATTAACACTAGCAAAAGTACAGCAAACTTCGTTAAAGAAACGACGAAGCCCAACATCA	217
Qy	256	ACCTATGATACATAT 270	
Db	216	ACCTATGATACATAT 202	

RESULT 9		
166614	486 bp	DNA
LOCUS	166614	linear
DEFINITION	Sequence 1 from patent US 5670623.	PAT 29-DEC-1997
FEATURES		

	/organism="unknown"			
BASE COUNT	194 a	89 c	79 g	124 t
ORIGIN				

Query Match	33.2%;	Score	255;	DB	6;	Length	486;
Best Local Similarity	100.0%;	Pred. No.	2	3e-139;			
Matches	255;	Conservative	0;	Mismatches	0;	Indels	0;
						Gaps	0;

Oy	16	TAAACCTGTAAACAATGAGCAAAACCAACGCAATTTACACCAATTAATAAAATTACTAC	75
Db	31	TACAACTTAAACAATGACGACAAACCAACTCAATTAACACCAATTAATCAAAATTACTAC	90
Oy	76	ACAATGTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATATTATTACACAGTGAT	135
Db	91	ACAATGTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATATTATTACACCAAGTGAT	150
Oy	136	GGTACACAGGACCAACCTTTCGTGCTGACCATGCTGTCATTTTGTAGAAATAGCTAT	195
Db	151	GSTACACAGGACCAAACTTTCGTGCTGACCATGCTGTCATTTATTGAAATATGCTAT	210
Oy	136	GTTGATTAACACTAGCAAAAGTGACGCAAACTTCGTTAAAGAAACGCAAGCCCAACATCA	255
Db	211	GTTGATTAACACTAGCAAAAGTGACGCAAACTTCGTTAAAGAAACGCAAGCCCAACATCA	270
Oy	256	ACCTATGATACATAT 270	
Db	271	ACCTATGATACATAT 285	

RESULT	10				
LOCUS	I66615/c				
DEFINITION	I66615	486 bp	DNA	linear	PAT 29-DEC-1997
ACCESSION	Sequence 3 from patent US 5670623.				
VERSION	I66615				
KEYWORDS	I66615.1 GI:2724593				
SOURCE	.				
ORGANISM	Unknown.				
REFERENCE	Unclassified:				
AUTHORS	1 (bases 1 to 486)				
TITLE	Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Dol,R.H.				
JOURNAL	Methods of use of cellulose binding domain proteins				
FEATURES	Patent: US 5670623-A 3 23-SEP-1997;				
	Location/Qualifiers				
	1..486				
	/organism="unknown"				
BASE COUNT	124 a	79 c	89 g	194 t	
ORIGIN					

Query Match	33.2%;	Score 255;	DB 6;	Length 486;
Best Local Similarity	100.0%;	Pred. No. 2,3e-159;		
Matches 255;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Oy	76	ACATCTGCAGTGGATTAAATTTAAATGACGTAAAGTTAGATATTTATTCACACAAGTAT	1359
Db	456	TACAACTCTAACCAAAATCGCACAACCAACACTCATTCACCCAAATAATCAAAATTACTAAC	397
Oy	396	ACATCTGCAGTGGATTAAATTTAAATGACGTAAAGTTAGATATTTATTCACACAAGTAT	337
Db	136	GSTACACAGAACAACCTTTTGTTGTTGACCATCTGCTGTCATTATTAGGAATAGCTAT	195
Oy	336	GSTACACAGAACAACCTTTTGTTGTTGACCATCTGCTGTCATTATTAGGAATAGCTAT	277
Db	196	GTTGATTACACGTAGCAAGTAGCAGCAAACTTCGTTAAGAAACAGACAGGCCAACATCA	258
Oy	276	GTTGATTACACTAGCAAAAGTAGCAGCAAACTTCGTTAAGAAACAGACAGGCCAACATCA	217
Db	256	ACCTATGATACATAT 270	
Oy	216	ACCTATGATACATAT 202	

RESULT	11
188789	
LOCUS	188789 486 bp DNA
DEFINITION	Sequence 1 from patent US 57,91044.
ACCESSION	188789
VERSION	188789.1 GI:3408729
KEYWORDS	.
	PAT 10-AUG-1998

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.
TITLE Cellulose binding domain fusion proteins
JOURNAL Patent: US 5719044-A 1 17-FEB-1998;
FEATURES Location/Qualifiers
source 1..486
BASE COUNT 194 a 89 c 79 g 124 t
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;
Best Local Similarity 100.0%; Pred. No. 2.3e-139;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGAGCACAACAACCTCAATTACACCAATTAATCAAAATTACTAAC 75
DB 31 TACAACCTTAACAATGAGCACAACAACCTCAATTACACCAATTAATCAAAATTACTAAC 90
QY 76 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 135
DB 91 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 150
QY 136 GGTCACACAAGGACAACTTCTGCTGTGACCATGCTGGTGATATTAGGAATAGCTAT 195
DB 151 GGTCACACAAGGACAACTTCTGCTGTGACCATGCTGGTGATATTAGGAATAGCTAT 210
QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTCTGTTAAAGAACAGCAAGCCCAACATCA 255
DB 211 GTTGATTAACACTAGCAAGTGACAGCAAACTCTGTTAAAGAACAGCAAGCCCAACATCA 270
QY 256 ACCTATGATACATAT 270
DB 271 ACCTATGATACATAT 285

RESULT 12
188790/c 188790 486 bp DNA linear PAT 10-AUG-1998
LOCUS Sequence 3 from patent US 5719044.
DEFINITION 188790
ACCESSION 188790.1 GI:3408730
VERSION
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 486)
AUTHORS Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.
TITLE Cellulose binding domain fusion proteins
JOURNAL Patent: US 5719044-A 3 17-FEB-1998;
FEATURES Location/Qualifiers
source 1..486
BASE COUNT 124 a 79 c 89 g 194 t
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;
Best Local Similarity 100.0%; Pred. No. 2.3e-139;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGAGCACAACAACCTCAATTACACCAATTAATCAAAATTACTAAC 75
DB 456 TACAACCTTAACAATGAGCACAACAACCTCAATTACACCAATTAATCAAAATTACTAAC 397
QY 76 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 135
DB 396 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 337
QY 136 GGTCACACAAGGACAACTTCTGCTGTGACCATGCTGGTGATATTAGGAATAGCTAT 195
DB 336 GGTCACACAAGGACAACTTCTGCTGTGACCATGCTGGTGATATTAGGAATAGCTAT 277

QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTCTGTTAAAGAACAGCAAGCCCAACATCA 255
DB 276 GTTGATTAACACTAGCAAGTGACAGCAAACTCTGTTAAAGAACAGCAAGCCCAACATCA 217
QY 256 ACCTATGATACATAT 270
DB 216 ACCTATGATACATAT 202

RESULT 13
AR096204 499 bp DNA linear PAT 08-SEP-2000
LOCUS Sequence 5 from patent US 6005092.
DEFINITION AR096204
ACCESSION AR096204
VERSION AR096204.1 GI:10024795
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 499)
AUTHORS Shoseyov,O. and Shani,Z.
TITLE Arabidopsis thaliana endo-1,4-.beta.-glucanase gene and promoter
JOURNAL Patent: US 6005092-A 5 21-DEC-1999;
FEATURES Location/Qualifiers
source 1..499
BASE COUNT 197 a 93 c 82 g 127 t
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 499;
Best Local Similarity 100.0%; Pred. No. 2.3e-139;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGAGCACAACAACCTCAATTACACCAATTAATCAAAATTACTAAC 75
DB 36 TACAACCTTAACAATGAGCACAACAACCTCAATTACACCAATTAATCAAAATTACTAAC 95
QY 76 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 135
DB 96 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 155
QY 136 GGTCACACAAGGACAACTTCTGCTGTGACCATGCTGGTGATATTAGGAATAGCTAT 195
DB 156 GGTCACACAAGGACAACTTCTGCTGTGACCATGCTGGTGATATTAGGAATAGCTAT 215
QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTCTGTTAAAGAACAGCAAGCCCAACATCA 255
DB 216 GTTGATTAACACTAGCAAGTGACAGCAAACTCTGTTAAAGAACAGCAAGCCCAACATCA 275
QY 256 ACCTATGATACATAT 270
DB 276 ACCTATGATACATAT 290

RESULT 14
AR151733 1146 bp DNA linear PAT 08-AUG-2001
LOCUS Sequence 21 from patent US 6232107.
DEFINITION AR151733
ACCESSION AR151733
VERSION AR151733.1 GI:15117783
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1146)
AUTHORS Bryan,B.J. and Szent-Gyorgyi,C.
TITLE Luciferases and fluorescent proteins and the use thereof in diagnostics, high throughput screening and novel items
JOURNAL Patent: US 6232107-A 21 15-MAY-2001;
FEATURES Location/Qualifiers
source 1..1146

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:56:01 : Search time 13.2642 Seconds
(without alignments)
471.414 Million cell updates/sec

Title: US-09-677-374-6

Perfect score: 1335
Sequence: 1 MSVEFYNSKSAQNTSITPI.....IYGTACPOPDGRMOVISTEK 256

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCPDUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	480.5	36.0	382	4	US-09-277-716-22 Sequence 22, Appl
2	477	35.7	162	1	US-08-048-164A-2 Sequence 2, Appl
3	477	35.7	162	1	US-08-460-462-2 Sequence 2, Appl
4	477	35.7	162	1	US-08-460-457-2 Sequence 2, Appl
5	477	35.7	162	1	US-08-460-458-2 Sequence 2, Appl
6	477	35.7	162	2	US-08-460-455-2 Sequence 2, Appl
7	477	35.7	162	2	US-08-330-394A-2 Sequence 2, Appl
8	477	35.7	163	3	US-09-006-636-7 Sequence 7, Appl
9	477	35.7	163	4	US-09-006-632-7 Sequence 7, Appl
10	472	35.4	154	2	US-08-330-394A-29 Sequence 29, Appl
11	472	35.4	156	2	US-08-330-394A-22 Sequence 22, Appl
12	245	18.4	167	5	PCT-US95-13813-9 Sequence 9, Appl
13	245	18.4	493	4	US-09-198-956-10 Sequence 10, Appl
14	245	18.4	493	4	US-09-198-955A-12 Sequence 12, Appl
15	149.5	11.2	1426	4	US-09-136-574A-43 Sequence 43, Appl
16	148.5	11.1	616	4	US-09-136-574A-47 Sequence 47, Appl
17	148.5	11.1	1751	4	US-09-136-574A-44 Sequence 44, Appl
18	122.5	9.2	551	2	US-09-033-537A-1 Sequence 1, Appl
19	100.5	7.5	700	2	US-07-862-588A-2 Sequence 2, Appl
20	85	6.4	1021	1	US-07-910-760-12 Sequence 12, Appl
21	85	6.4	1021	1	US-08-440-519-12 Sequence 12, Appl
22	85	6.4	1021	1	US-08-440-519-12 Sequence 12, Appl
23	83	6.2	382	4	US-08-444-818-68 Sequence 68, Appl
24	83	6.2	460	4	US-08-444-818-20 Sequence 20, Appl
25	83	6.2	512	4	US-08-867-611-58 Sequence 58, Appl
26	83	6.2	592	4	US-08-867-611-47 Sequence 47, Appl
27	83	6.2	594	4	US-08-867-611-48 Sequence 48, Appl

28	83	6.2	597	4	US-08-867-611-16 Sequence 16, Appl
29	83	6.2	597	5	PCT-US92-06965A-21 Sequence 21, Appl
30	83	6.2	599	4	US-08-867-611-18 Sequence 18, Appl
31	83	6.2	599	5	PCT-US92-06965A-23 Sequence 23, Appl
32	83	6.2	739	4	US-08-867-611-49 Sequence 49, Appl
33	83	6.2	739	4	US-08-444-818-148 Sequence 148, App
34	83	6.2	781	4	US-08-867-611-4 Sequence 4, Appl
35	83	6.2	781	5	PCT-US92-06965A-9 Sequence 9, Appl
36	83	6.2	859	4	US-08-444-818-30 Sequence 30, Appl
37	83	6.2	970	1	US-08-375-709-7 Sequence 7, Appl
38	83	6.2	970	1	US-08-752-929-7 Sequence 7, Appl
39	83	6.2	970	4	US-09-090-793-5 Sequence 5, Appl
40	83	6.2	971	4	US-08-867-611-52 Sequence 52, Appl
41	83	6.2	971	4	US-08-867-611-53 Sequence 53, Appl
42	83	6.2	992	4	US-08-867-611-54 Sequence 54, Appl
43	83	6.2	1786	4	US-08-444-818-54 Sequence 54, Appl
44	83	6.2	2261	4	US-08-444-818-66 Sequence 66, Appl
45	83	6.2	2772	4	US-08-444-818-89 Sequence 89, Appl

ALIGNMENTS

```
RESULT 1
US-09-277-716-22
; Sequence 22, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,710A
; EARLIER FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 22
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Artificial Sequence: fusion protein
; FEATURE:
; OTHER INFORMATION: Cellulose Binding Domain (CBD)-Gaussia luciferase fusion prote
US-09-277-716-22
Query Match          36.0%; Score 480.5; DB 4; Length 382;
Best Local Similarity 62.6%; Pred. No. 4.1e-41;
Matches 107; Conservative 10; Mismatches 33; Indels 21; Gaps 4;
OY 1 MSVEFYNSKSAQNTSITPIIKITNTSDSLNLDVAVRYRYTSDGOGCFWCDHMGAL 60
Db 1 MSVEFYNSKSAQNTSITPIIKITNTSDSLNLDVAVRYRYTSDGOGCFWCDHMGAL 60
OY 61 LGNSYVNTSKVTANFKEFASPTSTYDTYLDPSHMRCL---CGSSLITISFLVGCQAQ 117
Db 61 LGNSYVNTSKVTANFKEFASPTSTYDTYLDPSHMRCL---CGSSLITISFLVGCQAQ 117
OY 118 NESRQ----EVGAATGAIVGCVAGQLFGKSGRVSNAIGCA-VLGLIGS 162
Db 121 NYTQINDYSFSSSTFVNPKVTGY-----IGGAKVLTGAPGS 159
RESULT 2
US-08-048-164A-2
; Sequence 2, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
```


APPLICANT: DOI, ROY H.
TITLE OF INVENTION: METHODS OF DETECTION USING THE
CELLULOSE BINDING DOMAIN
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/330,394A
FILING DATE: 27-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CIP OF PCT/US94/04132
FILING DATE: 14-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REFERENCE/DOCKET NUMBER: 7809-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66441 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-330-394A-2

Query Match 35.7%; Score 477; DB 2; length 162;
Best Local Similarity 97.8%; Pred. No. 2.7e-41;
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYYTSDGTGGTFCMDHAGAL 60
|||||
DB 6 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYYTSDGTGGTFCMDHAGAL 65
|||||

QY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYLD 92
|||||

DB 66 LGNSYVDNTSKVTANFVKETASPTSTYDYLD 97
|||||

RESULT 8
US-09-006-636-7
Sequence 7, Application US/09006636
Patent No. 6005092
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shani, Ziv
TITLE OF INVENTION: ARABIDOPSIS THALIANA ENDO-1,4-BETA-
GLUCANASE GENE, PROMOTER AND PROTEIN
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/006,636
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7809-019
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 163 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-006-636-7

Query Match 35.7%; Score 477; DB 3; length 163;
Best Local Similarity 97.8%; Pred. No. 2.7e-41;
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYYTSDGTGGTFCMDHAGAL 60
|||||

DB 7 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYYTSDGTGGTFCMDHAGAL 66
|||||

QY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYLD 92
|||||

DB 67 LGNSYVDNTSKVTANFVKETASPTSTYDYLD 98
|||||

RESULT 9
US-09-006-632-7
Sequence 7, Application US/09006632
Patent No. 618440
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shani, Ziv
TITLE OF INVENTION: TRANSGENIC PLANTS OF ALTERED
MORPHOLOGY
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,632
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7809-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 163 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-006-632-7

Query Match 35.7%; Score 477; DB 4; Length 163;
Best Local Similarity 97.8%; Pred. No. 2.7e-41;
Matches 90; Conservative 2; Mismatches 0; Indels 0; Caps 0;

QY 1 MSVEFYNSKSAQNTSTPIIKITNTSDSLNDVKKVRYTSDGTGQGFMCDAAGAL 60
DB 7 MSVEFYNSKSAQNTSTPIIKITNTSDSLNDVKKVRYTSDGTGQGFMCDAAGAL 66
QY 61 GNSYVDNTSKVTANFVKETASPTSTYDTYLD 92
DB 67 GNSYVDNTSKVTANFVKETASPTSTYDTYVE 98

RESULT 10

US-08-330-394A-29
Sequence 29, Application US/08330394A
Patent No. 5856201

GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Yosef, Karmey
APPLICANT: Shpiegl, Itai
APPLICANT: Goldstei, Marc A.
APPLICANT: Dol, Roy H.
TITLE OF INVENTION: METHODS OF DETECTION USING THE
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/330,394A
FILING DATE: 27-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CIP OF PCT/US94/04132
FILING DATE: 14-APR-1994

ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REFERENCE/DOCKET NUMBER: 7809-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66441 PENNIE

INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 154 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-330-394A-29

Query Match 35.4%; Score 472; DB 2; Length 154;
Best Local Similarity 97.8%; Pred. No. 8.2e-41;
Matches 89; Conservative 2; Mismatches 0; Indels 0; Caps 0;

QY 2 SVEFYNSKSAQNTSTPIIKITNTSDSLNDVKKVRYTSDGTGQGFMCDAAGAL 61
DB 1 SVEFYNSKSAQNTSTPIIKITNTSDSLNDVKKVRYTSDGTGQGFMCDAAGAL 60
QY 62 GNSYVDNTSKVTANFVKETASPTSTYDTYLD 92

DB 61 GNSYVDNTSKVTANFVKETASPTSTYDTYVE 91

RESULT 11

US-08-330-394A-22
Sequence 22, Application US/08330394A
Patent No. 5856201

GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Yosef, Karmey
APPLICANT: Shpiegl, Itai
APPLICANT: Goldstei, Marc A.
APPLICANT: Dol, Roy H.
TITLE OF INVENTION: METHODS OF DETECTION USING THE
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/330,394A
FILING DATE: 27-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CIP OF PCT/US94/04132
FILING DATE: 14-APR-1994

ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REFERENCE/DOCKET NUMBER: 7809-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66441 PENNIE

INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-330-394A-22

Query Match 35.4%; Score 472; DB 2; Length 156;
Best Local Similarity 97.8%; Pred. No. 8.4e-41;
Matches 89; Conservative 2; Mismatches 0; Indels 0; Caps 0;

QY 2 SVEFYNSKSAQNTSTPIIKITNTSDSLNDVKKVRYTSDGTGQGFMCDAAGAL 61
DB 1 SVEFYNSKSAQNTSTPIIKITNTSDSLNDVKKVRYTSDGTGQGFMCDAAGAL 60
QY 62 GNSYVDNTSKVTANFVKETASPTSTYDTYLD 92
DB 61 GNSYVDNTSKVTANFVKETASPTSTYDTYVE 91

RESULT 12

PCT-US95-13813-9
Sequence 9, Application PC/RUS9513813

GENERAL INFORMATION:
APPLICANT: Yeda Research and Development Co. Ltd.
APPLICANT: Ramot University Authority for Applied
Research and Industrial Development Ltd.
APPLICANT: Technion Research and Development Foundation Ltd.
APPLICANT: Bayer, Edward A.
APPLICANT: Morag, Ely
APPLICANT: Wilchek, Melf

```

APPLICANT: Lamed, Raphael
APPLICANT: Shoham, Yuval
TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
TITLE OF INVENTION: PROTEINS AND USE THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Broadway and Nelmark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13813
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: BAYER-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO.: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 167 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-13813-9

Query Match      18.4%; Score 245; DB 5; Length 167;
Best Local Similarity 51.5%; Pred. No. 1,7e-17;
Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;

QY      1 MSVEFYNSKNAQTNSITFLIKITNTSDSLNDVKRYVYTSGTGOTWCHAGAL 60
        : ||||| | | | | | | | | | | | | | | | | | | | | | | | |
DB      9 LAKEFINSNSDPTNINSIPQFKVTNTGSSAIDLSKLTLRYVTDOKDOTWCDA-AI 67
        : | : | | | | | | | | | | | | | | | | | | | | | | | |
QY      61 LGN-STVDNTSKYTANFEKETASPSYSTDYTDPSHMGGLD 101
        : | : | | | | | | | | | | | | | | | | | | | | | | | |
DB      68 IGSNGSYNGITSNVKGTFEVMSSS-TNNADTYLIEISFTGGTLE 109
        : | : | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-09-198-956-10
: Sequence 10, Application US/09198956
: Patent No. 6165769
: GENERAL INFORMATION:
: APPLICANT: Andersen, Lene N.
: APPLICANT: Schuelein, Martin
: APPLICANT: Lange, Niels Erik K.
: APPLICANT: Bjornvad, Mads E.
: APPLICANT: Schmitt, Kirk
: TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
: FILE REFERENCE: 5377.200-US
: CURRENT APPLICATION NUMBER: US/09/198,956
: CURRENT FILING DATE: 1998-11-24
: EARLIER FILING DATE: 1997-11-24
: EARLIER APPLICATION NUMBER: 60/067,240
: EARLIER FILING DATE: 1997-12-02
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 10

```

```

; LENGTH: 493
; TYPE: PROT
; ORGANISM: Bacillus licheniformis
US-09-198-956-10

Query Match          18.4%; Score 245; DB 4; Length 493;
Best Local Similarity 51.5%; Pred. No. 8.3e-17;
Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;

OY      1 MSVEFYNSNKAQNSTPIPIKINTSDSLNLDVYVRRYYYSDDGTGOTFWCCHAGAL 60
        : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      335 LKVEFYNSNPEDTJNSINPOKFKYNTGSSAIDLSKLTFRYYVYDGGKDTFWCDBA-AI 393
        : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      61 LGN--SYVDNTSKVTANFVKETASPTSTYTDYLDPSHMRGLO 101
        : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      394 IGSNGSYNGITSNVKGFVKKMSS--TNNADTYLEISFTGTLLE 435
        : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: NO. 6187580e1 Pectate Lyases
; FILE REFERENCE: 53/8.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PROT
; ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match          18.4%; Score 245; DB 4; Length 493;
Best Local Similarity 51.5%; Pred. No. 8.3e-17;
Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;

OY      1 MSVEFYNSNKAQNSTPIPIKINTSDSLNLDVYVRRYYYSDDGTGOTFWCCHAGAL 60
        : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      335 LKVEFYNSNPEDTJNSINPOKFKYNTGSSAIDLSKLTFRYYVYDGGKDTFWCDBA-AI 393
        : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      61 LGN--SYVDNTSKVTANFVKETASPTSTYTDYLDPSHMRGLO 101
        : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      394 IGSNGSYNGITSNVKGFVKKMSS--TNNADTYLEISFTGTLLE 435
        : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366

```

```

1 GENERAL INFORMATION:
2
3 APPLICANT: Farrington, Graham K.
4
5 Anderson, Paige
6 Gibbs, Moreland
7 Bergquist, Peter
8 Daniels, Roy
9 Morgan, Hugh W.
10 Williams, Diane P.
11
12 TITLE OF INVENTION: Compositions and Methods for
13 Treating Cellulose Containing Fabrics Using Truncated
14 Cellulase Enzyme Compositions
15
16 NUMBER OF SEQUENCES: 49
17
18 CORRESPONDENCE ADDRESS:
19 ADDRESSEE: Howson and Howson
20 STREET: Spring House Corporate Center, P.O. Box 457
21 CITY: Spring House
22 STATE: PA
23 COUNTRY: USA
24 ZIP: 19477
25
26 COMPUTER READABLE FORM:
27
28 MEDIUM TYPE: Diskette
29 COMPUTER: IBM Compatible
30 OPERATING SYSTEM: DOS
31
32 CURRENT APPLICATION DATA:
33 APPLICATION NUMBER: US/09/136,574A
34 FILING DATE: 19-Aug-1998
35
36 CLASSIFICATION: <unknown>
37
38 PRIOR APPLICATION DATA:
39 APPLICATION NUMBER: US 08/932,571
40 FILING DATE: September 19, 1997
41
42 ATTORNEY/AGENT INFORMATION:
43 NAME: Bak, Mary E.
44 REGISTRATION NUMBER: 31,215
45 REFERENCE/DOCKET NUMBER: 1997US001/CIP
46
47 TELECOMMUNICATION INFORMATION:
48 TELEPHONE: 215-540-9200
49 TELEFAX: 215-540-5818
50
51 TELEX: <unknown>
52
53 INFORMATION FOR SEQ ID NO: 43:
54
55 SEQUENCE CHARACTERISTICS:
56 LENGTH: 1426 amino acids
57 TYPE: amino acid
58 STRANDEDNESS: single
59 TOPOLOGY: linear
60
61 MOLECULE TYPE: No. 6294366e
62
63 SEQUENCE DESCRIPTION: SEQ ID NO: 43:
64
65 US-09-136-574A-43
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
```

Search completed: October 27, 2002, 11:00:30
Job time : 15.2642 secs

THIS PAGE BLANK (USPTO)

	12	477	35.7	152	20	AAM90077	C. cellulovorans C.
	13	472	35.4	156	20	AAM90081	C. cellulovorans C.
	14	472	35.4	156	20	AAM90080	C. cellulovorans C.
	15	245	18.4	167	17	AAR95080	Cellulose binding
	16	245	18.4	167	21	AAV54123	A mannanase-linked
	17	245	18.4	493	20	AAV28650	Pectate lyase-Limn
	18	245	18.4	493	20	AAV43318	Pectate lyase CBD
	19	245	18.4	531	18	AAM15338	scaffoldin protein
	20	245	18.4	1853	19	AAM43108	C. thermocellum c
	21	149.5	11.2	1426	20	AAV13492	Truncated cellulase
	22	148.5	11.1	616	20	AAV13494	Truncated cellulase
	23	148.5	11.1	1751	20	AAV13493	Truncated cellulase
	24	127.5	9.6	1352	22	AAG63962	Amino acid sequen
	25	126.5	9.5	1350	22	AAG63963	Amino acid sequen
	26	122.5	9.2	551	18	AAM18790	Corrected Bacillus
	27	113.5	8.5	499	14	AAK42122	NK-1 cellulase.
	28	112	8.4	20	22	AAB81130	Ospa B-cell epitop
	29	111	8.3	224	22	AAB20105	Moraxella catarrh
	30	102.5	7.7	223	20	AAV34487	Porphyromonas gin
	31	102.5	7.7	220	20	AAV34362	Porphyromonas gin
	32	102	7.6	782	12	AAE15625	Cellulase AE-1
	33	101.5	7.6	154	11	AAR05799	PBOMP-2 gene prod
	34	100.5	7.5	700	12	AAE13327	Novel endoglucanase
	35	98	7.3	922	20	AAV34597	Chlamydia pneumon
	36	98	7.3	922	20	AAM89419	Chlamydia pneumon
	37	98	7.3	922	21	AAI95548	Chlamydia pneumon
	38	95	7.1	309	22	ABG15906	Novel human diagn
	39	95	7.1	2309	22	ABB66232	Drosophila melano
	40	92	6.9	285	20	AAV41688	Human PRO284 prot
	41	92	6.9	285	21	AAH44244	Human PRO284 (UNQ
	42	92	6.9	285	22	AAU29025	Human PRO polypept
	43	91.5	6.9	285	21	AAH87280	Human signal pept
	44	91.5	6.9	285	21	AAV73440	Human secreted pro
	45	91.5	6.9	285	22	AAO93388	Human protein HP1

ALIGNMENTS

RESULT_1	
AAAB81128	
ID	AAAB81128 standard; Protein; 256 AA.
XX	
AC	AAAB81128;
XX	
DT	11-JUL-2001 (first entry)
DE	
CC	C17E2 OSPA construct with N-terminal fusion partner.
XX	
KM	Polkiothermic fish; piscirickettsia salmonis; rickettsial pathogen;
KW	vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
KW	SRS; 17E2; fusion construct.
XX	
OS	Piscirickettsia salmonis.
OS	Synthetic.
XX	
Key	Location/Qualifiers
FH	1..95
FT	/label= undefined_N-terminal_fusion_partner
FT	Region
FT	96..256
FT	/label= C17E2_OSPA
FT	/note="Product of Ospa gene optimised for expression in Escherichia coli"
XX	
PN	CA2281913-A1.
XX	
PD	17-MAR-2001.
XX	
PF	17-SEP-1999; 99CA-2281913.
XX	
PR	17-SEP-1999; 99CA-2281913.
PA	(KAYW/) KAY W W.

PA (BURJ/) BURIAN J.
 PA (KUZU/) KUZUK M A.
 PI Kay WW, Burian J, Kuzuk MA;
 XX WPI: 2001-316844/34.
 DR N-PSDB; AAF86248.
 PT Method for protecting polikiothermic fish against salmonid rickettsial
 PT septicemia and other rickettsial diseases comprises administering a
 PT vaccine containing the ospa protein of Piscirickettsia salmonis
 XX
 PS Example 4; Fig 5; 35pp; English.
 XX This invention relates to a method for the protection against infection
 CC of a polikiothermic fish by the bacterial pathogen, Piscirickettsia
 CC salmonis. The method comprises administering an immunogenic amount of a
 CC P. salmonis specific antigen termed OSPA, or an immunogenic fragment of
 CC OSPA in the form of a vaccine. The method is used for protecting animals,
 CC particularly polikiothermic fish, against the bacterial pathogen
 CC P. salmonis. The method is also useful for protecting against salmonid
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
 CC sequence represents the amino acid sequence of C17E2, a P. salmonis OSPA
 CC construct optimised for expression in Escherichia coli, fused to an
 CC undefined N-terminal fusion partner. The fusion protein is used in a
 CC vaccine to create an anti-OSPA antibody response.
 CC
 SQ Sequence 256 AA:
 Query Match 100.0%; Score 1335; DB 22; Length 256;
 Best Local Similarity 100.0%; Pred. No. 1.5e-122;
 Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSVEFYNSKSKSQSTIPITIKITNTSDNLNDVKRYIYTSIDGTOGTFMCDHAGAL 60
 DB 1 MSVEFYNSKSKSQSTIPITIKITNTSDNLNDVKRYIYTSIDGTOGTFMCDHAGAL 60
 QY 61 LGNSYDNTSKVTANFVKETASPTSTYDYLDPMSHRCLOGSSLIITISVFLVCAONFS 120
 DB 61 LGNSYDNTSKVTANFVKETASPTSTYDYLDPMSHRCLOGSSLIITISVFLVCAONFS 120
 QY 121 ROEVAATGAVVAGVAGOLFPGKSGRYSMAIGAVLGLIGSKIGOSMDQDKIKLNSL 180
 DB 121 ROEVAATGAVVAGVAGOLFPGKSGRYSMAIGAVLGLIGSKIGOSMDQDKIKLNSL 180
 QY 181 EKKVAGVTRMRNPDGNSYSVEPVRYTORYNKQERRQOYCEPQOKAMIAQOKETIGT 240
 DB 181 EKKVAGVTRMRNPDGNSYSVEPVRYTORYNKQERRQOYCEPQOKAMIAQOKETIGT 240
 QY 241 ACPOPGRMOYISTEK 256
 DB 241 ACPOPGRMOYISTEK 256
 RESULT 2
 ID AAB81127 standard; Protein; 161 AA.
 AC AAB81127;
 XX 11-JUL-2001 (first entry)
 DE Optimised Ospa protein 17E2 amino acid sequence.
 XX Polikiothermic fish; piscirickettsia salmonis; rickettsial pathogen;
 KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;
 KW SRS.
 OS Piscirickettsia salmonis.
 OS Synthetic.
 XX Key Location/Qualifiers
 FH Region 109..128
 FT

FT
 XX CA2281913-A1. /label= B_cell_epitope
 PN 17-MAR-2001.
 PD 17-SEP-1999; 99CA-2281913.
 XX 17-SEP-1999; 99CA-2281913.
 PF 17-SEP-1999; 99CA-2281913.
 PR 17-SEP-1999; 99CA-2281913.
 XX
 PA (KAYW/) KAY W W.
 PA (BURJ/) BURIAN J.
 PA (KUZU/) KUZUK M A.
 PI Kay WW, Burian J, Kuzuk MA;
 XX WPI: 2001-316844/34.
 DR N-PSDB; AAF86247.
 PT Method for protecting polikiothermic fish against salmonid rickettsial
 PT septicemia and other rickettsial diseases comprises administering a
 PT vaccine containing the ospa protein of Piscirickettsia salmonis
 XX
 PS Disclosure; Fig 5; 35pp; English.
 XX This invention relates to a method for the protection against infection
 CC of a polikiothermic fish by the bacterial pathogen, Piscirickettsia
 CC salmonis. The method comprises administering an immunogenic amount of a
 CC P. salmonis specific antigen termed OSPA, or an immunogenic fragment of
 CC OSPA in the form of a vaccine. The method is used for protecting animals,
 CC particularly polikiothermic fish, against the bacterial pathogen
 CC P. salmonis. The method is also useful for protecting against salmonid
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
 CC sequence represents optimised P. salmonis OSPA protein 17E2. The DNA
 CC encoding Ospa 17E2 (AAF86247) has been optimised for expression in
 CC Escherichia coli. An Ospa protein with an N-terminal fusion partner is
 CC used in a vaccine to create an anti-OSPA antibody response.
 CC
 SQ Sequence 161 AA:
 Query Match 62.3%; Score 832; DB 22; Length 161;
 Best Local Similarity 100.0%; Pred. No. 1.6e-73;
 Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 96 MRGCLGSSLIITISVFLVCAONFSROEYGAATGAVVAGVAGOLFPGKSGRYSMAIGAV 155
 DB 1 MRGCLGSSLIITISVFLVCAONFSROEYGAATGAVVAGVAGOLFPGKSGRYSMAIGAV 60
 QY 156 LGGLIGSKIGOSMDQDKIKLNSLEKVRAGVYTRMRNPDGNSYSVEPVRYTORYNKOE 215
 DB 156 LGGLIGSKIGOSMDQDKIKLNSLEKVRAGVYTRMRNPDGNSYSVEPVRYTORYNKOE 120
 QY 216 RROQYCREFOOKAMIAQOKETIGTACPOPGRMOYISTEK 256
 DB 121 RROQYCREFOOKAMIAQOKETIGTACPOPGRMOYISTEK 161
 RESULT 3
 ID AAG78025 standard; Protein; 162 AA.
 AC AAG78025;
 XX 15-JAN-2002 (first entry)
 DE Piscirickettsia salmonis polypeptide p10.6.
 XX Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
 KW septicemia; SRS; surface antigen; vaccine; antibacterial; fish;
 KW ATCC VR-1361.
 OS Piscirickettsia salmonis.
 OS

XX 17-SEP-1999; 99CA-2281913.
XX
XX
PR 17-SEP-1999; 99CA-2281913.
XX
XX
PA (KAYW/) KAY M M.
PA (BURJ/) BURIAN J.
PA (KUZV/) KUZYK M A.
XX
XX
PL Kay WW, Burian J, Kuzyk MA;
XX
XX WPI: 2001-316844/34.
DR
DR N-PSDB; AAF86246.
XX
XX
XX Method for protecting polkithohermic fish against salmonid rickettsial
PT septicemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of Piscirickettsia salmonis
XX
XX
XX Example 2: Fig 2B; 35pp: English.
XX
XX
XX This invention relates to a method for the protection against infection
CC of a polkithohermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkithohermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
CC sequence represents P. salmonis Ospa protein. An Ospa protein with an
CC N-terminal fusion partner is used in a vaccine to create an anti-Ospa
CC antibody response.
XX
XX
XX Sequence 162 AA:
XX

	Query Match	61.0%;	Score 815;	DB 22;	Length 162;	
	Best Local Similarity	98.8%;	Pred. No. 7.6e-72;			
	Matches 158; Conservative	1;	Mismatches	1;	Indels 0;	Gaps
Oy	97 RCLGSSSLITISVPLVGCQNFNSRQEVGAATGAVVGAGOLFEGKSGRYSMAGAVL	156				
Db	3 RCLGSSSLITISVPLVGCQNFNSRQEVGAATGAVVGAGOLFEGKSGRYSMAGAVL	62				
Oy	157 GGLIGSKICQSDQDQIKILNQSLEKKAGQVTRMRPDTGNSTSVPRYYQRNKQER	216				
Db	63 GGLIGSKICQSDQDQIKILNQSLEKKAGQVTRMRPDTGNSTSVPRYYQRNKQER	122				
Oy	217 ROOYCREFOOKAMIIAGOKOEIYGTAPOPDPGRMOVISTEK	256				
Db	123 ROOYCREFOOKAMIIAGOKOEIYGTAPOPDPGRMOVISTEK	162				
RESULT 5						
AAE05746						
ID AAE05746 standard; Protein; 190 AA.						
XX	AAE05746;					
AC						
XX						
DT	24-SEP-2001 (first entry)					
DE	Clostridium cellulovorans cellulose binding domain-180 (CBD-180).					
XX						
KW	Polysaccharide modification; polysaccharide binding domain; PBD; paper;					
KW	year; fiber; textile; biological crosslinker; mechanical property;					
KM	Clostridium cellulovorans cellulose binding domain; wet strength;					
KW	durability; elasticity; CBDelos; cellulose binding protein A; CBP A;					
KM	CBD-180.					
XX						
OS	Clostridium cellulovorans.					
XX						
PN	WO200134091-A2.					
PD	17-MAY-2001.					
XX						

```

PF 02-NOV-2000; 2000MO-IL00708.
XX 08-NOV-1999; 99US-0164140.
XX 18-NOV-1999; 99US-0166389.
XX (CBPT-) CBD TECHNOLOGIES LTD.
PA (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX Levy I, Nussinovitch A, Shoseyov O;
PI MPI: 2001-457121/49.
DR N-PSDB; AAD11043.
XX Preparation of a polysaccharide containing material having at least one
PT desired structural, chemical, physical, electrical and/or mechanical
PT property -
XX Example 1.2; Fig 1e-1g; 121pp; English.
PS The present invention relates to methods and compositions for cross-
XX linking and/or modifying the properties of polysaccharide materials.
CC The method involves treating the polysaccharide structure with a
CC polysaccharide binding domain (PBD) fusion protein. The method is
CC used to alter the structural, chemical, physical, electrical and
CC mechanical properties of polysaccharide materials such as paper,
CC yarns, fibers and textiles, using biological crosslinking agents.
CC The polysaccharide containing materials have improved mechanical
CC properties such as wet strengths, durability and elasticity. The PBD
CC reagent is applied in the sizing step. The use of a biological crosslinker
CC which eliminates the sizing step. The use of a biological crosslinker
CC improves the recyclability of paper products. The PBD reagent maintains
CC the fine fibers in a slurry therefore resulting in better recovery of
CC raw materials. The PBD molecules are eluted by strong alkaline conditions
CC which enhances the ability of the alkaline glue used in binding to
CC penetrate paper.
CC The present sequence is Clostridium cellulovorans cellulose binding
CC domain (CBD)-180. CBD-180 contains 180 amino acids of C. cellulovorans
CC CBD (CBDC180) of cellulose binding protein A (CBP A).
XX SQ
Sequence 190 AA:
Query Match 36.5%; Score 487; DB 22; Length 190;
Best Local Similarity 46.8%; Pred. No. 1.4e-39;
Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;
OY 1 MSVEFYNSKSAQTNSITPIIKITNTSDSLNDLVKRYRYTSDGTGQTFWCDHAGAL 60
DB 1 MSVEFYNSKSAQTNSITPIIKITNTSDSLNDLVKRYRYTSDGTGQTFWCDHAGAL 60
OY 61 LGNSVYDNTSKYTANFVETASPTSTYDTPDSHMRGCLQSSLIITSVLWCAOMFS 120
DB 61 LGNSVYDNTSKYTANFVETASPTSTYDTPDSHMRGCLQSSLIITSVLWCAOMFS 120
OY 121 ROEVGAATGAVGVAGOLFEGKSGRVSMAIGAVLGLIGSKIQGSDQDQKIKLNSL 180
DB 93 -----EGFASGRATL-----KKQSFITIQ----- 111
OY 181 EKYKAGQVTR--WRNPDTGNSYSVEPRYRYQYKNOERHQYCRFEOQKAMLAGOKET 238
DB 112 -----GRITKSDMSNTYTQNDYSFDSASTPYVNP-----KYTGITGAK--VL 153
OY 239 GTACPOPD 246
DB 154 GTR-PCPD 160

```

```

XX DE Chimeric S peptide-cellulose binding domain-S protein.
XX XX Polysaccharide modification; polysaccharide binding domain; PBD; paper;
XX KW yarn; fiber; textile; biological crosslinker; mechanical property;
XX KW wet strength; durability; elasticity; cellulose binding domain; CBD;
XX KW chimeric protein; S peptide-cellulose binding domain-S protein;
XX KW Slep-CBD-Sprot; bovine.
XX OS Chimeric - Clostridium cellulovorans.
XX OS Chimeric - Bos sp.
XX Key Location/Qualifiers
XX FH Region 30..208
XX FT /note="This region is derived from C. cellulovorans"
XX FT Region 226..326
XX FT /note="This region is derived from bovine"
XX FT Misc-difference 327
XX FT /label="Unknown"
XX FT /note="Encoded by TAG"
XX FT Misc-difference 327..328
XX FT /note="These residues are absent in the sequence shown
XX as SEQ ID NO: 10 in the sequence listing"
XX PF WO200134091-A2.
XX PD 17-MAY-2001.
XX PF 02-NOV-2000; 2000MO-IL00708.
XX 08-NOV-1999; 99US-0164140.
XX PR 18-NOV-1999; 99US-0166389.
XX XX (CBPT-) CBD TECHNOLOGIES LTD.
XX PA (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX PI Levy I, Nussinovitch A, Shoseyov O;
XX DR N-PSDB; AAD11046.
XX Preparation of a polysaccharide containing material having at least one
XX PT desired structural, chemical, physical, electrical and/or mechanical
XX PT property -
XX PS Example 4; Fig 4b-4g; 121pp; English.
XX XX The present invention relates to methods and compositions for cross-
XX linking and/or modifying the properties of polysaccharide materials.
XX CC The method involves treating the polysaccharide structure with a
XX polysaccharide binding domain (PBD) fusion protein. The method is
XX used to alter the structural, chemical, physical, electrical and
XX mechanical properties of polysaccharide materials such as paper,
XX yarns, fibers and textiles, using biological crosslinking agents.
XX CC The polysaccharide containing materials have improved mechanical
XX properties such as wet strengths, durability and elasticity. The PBD
XX reagent is applied in the sizing step. The use of a biological crosslinker
XX which eliminates the sizing step. The use of a biological crosslinker
XX improves the recyclability of paper products. The PBD reagent maintains
XX the fine fibers in a slurry therefore resulting in better recovery of
XX raw materials. The PBD molecules are eluted by strong alkaline conditions
XX which enhances the ability of the alkaline glue used in binding to
XX penetrate paper.
XX CC The present sequence is S peptide-cellulose binding domain-S protein
XX (Slep-CBD-Sprot), a fusion protein derived from Clostridium cellulovorans
XX and bovine.
XX Note: This sequence is stated to be the same as that shown as
XX SEQ ID NO: 10 in the sequence listing of the specification. However
XX this sequence has 2 additional residues at its C-terminal.
XX SQ Sequence 328 AA:
Query Match 36.5%; Score 487; DB 22; Length 328;

```

Best Local Similarity 46.8%; Pred. No. 2.9e-39;
Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;

```

OY 1 MSVEFYNSKNSAQTNSTPTPIKITTNSDSDLNLDVAVRRYRYSYSDGQGTFCMDHAGAL 60
DB 35 MSVEFYNSKNSAQTNSTPTPIKITTNSDSDLNLDVAVRRYRYSYSDGQGTFCMDHAGAL 94
OY 61 LGNSYVDNTSKVTANFVKETASPTSTYDTYLDPSHMRGCLQGSLLIISVFLVCAQNF 120
DB 95 LGNSYVDNTSKVTANFVKETASPTSTYDTYLDPSHMRGCLQGSLLIISVFLVCAQNF 126
OY 121 ROEYGAATGAVGAGVAGQLFGKSGSRVSMAGAVLGLGSKIGSQMDQDKIKLNQSL 180
DB 127 -----FGFASGRATL-----KKGQFTITQ----- 145
OY 181 EYKAKAGVTR--WRNPDTGNSYSVEPRYTORYNKORRROOYCREFOKAMIAQOKEIY 238
DB 146 ----GRITKSDWSNNTOTNDYSFDASSSTPVVNP-----KVTGYIGAK--VL 187
OY 239 GTACPOPD 246
DB 188 GTA-PCPD 194

```

RESULT 7
ID AAE05747
AAE05747 standard; protein; 341 AA.

AC AAE05747;

DT 24-SEP-2001 (first entry)

DE Clostridium cellulovorans CBD cross linker protein (CCP).

KM Polysaccharide modification; polysaccharide binding domain; PBD; paper;
KM yarn; fiber; textile; biological crosslinker; mechanical property;
KM wet strength; durability; elasticity; cellulose binding domain; CBD;

KM CBD cross linker protein; CCP.

OS Clostridium cellulovorans.

XX Key Location/Qualifiers

FT Misc-difference 340..341

FT /note- "Encoded by CCATAGAT"

FT Misc-difference 341

FT /note- "This residue is absent in the sequence shown as SEQ ID NO: 6 in the sequence listing"

PN W0200134091-A2.

PD 17-MAY-2001.

XX 02-NOV-2000; 2000MO-IL00708.

XX 08-NOV-1999; 99US-0164140.

PR 18-NOV-1999; 99US-0166389.

XX (CBDT-) CBD TECHNOLOGIES LTD.

PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.

XX Levy I, Nussinovitch A, Shoseyov O;

PI WPI: 2001-457121/49.

DR N-PSDB; AAD11044.

XX Preparation of a polysaccharide containing material having at least one

PT desired structural, chemical, physical, electrical and/or mechanical

XX property

PS Example 1.2; Fig 2b-2e; 121pp; English.

XX The present invention relates to methods and compositions for cross-

CC linking and/or modifying the properties of polysaccharide materials.

CC The method involves treating the polysaccharide structure with a
CC polysaccharide binding domain (PBD) fusion protein. The method is
CC used to alter the structural, chemical, physical, electrical and
CC mechanical properties of polysaccharide materials such as paper,
CC yarns, fibers and textiles, using biological crosslinking agents.
CC The polysaccharide containing materials have improved mechanical
CC properties such as wet strengths, durability and elasticity. The PBD
CC reagent is applied in the forming stage in fluting paper manufacture
CC which eliminates the sizing step. The use of a biological crosslinker
CC improves the recyclability of paper products. The PBD reagent maintains
CC the fine fibers in a slurry therefore resulting in better recovery of
CC raw materials. The PBD molecules are eluted by strong alkaline conditions
CC which enhances the ability of the alkaline glue used in binding to
CC penetrate paper.

CC The present sequence is Clostridium cellulovorans cellulose binding

CC domain (CBD) cross linker protein (CCP).

CC Note: This sequence is stated to be the same as that shown as

CC SEQ ID NO: 6 in the sequence listing of the specification. However

CC this sequence has an additional residue at its C-terminal.

CC

XX

SO Sequence 341 AA:

Query Match 36.5%; Score 487; DB 22; Length 341;

Best Local Similarity 46.8%; Pred. No. 3.1e-39;

Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;

```

OY 1 MSVEFYNSKNSAQTNSTPTPIKITTNSDSDLNLDVAVRRYRYSYSDGQGTFCMDHAGAL 60
DB 1 MSVEFYNSKNSAQTNSTPTPIKITTNSDSDLNLDVAVRRYRYSYSDGQGTFCMDHAGAL 60
OY 61 LGNSYVDNTSKVTANFVKETASPTSTYDTYLDPSHMRGCLQGSLLIISVFLVCAQNF 120
DB 61 LGNSYVDNTSKVTANFVKETASPTSTYDTYLDPSHMRGCLQGSLLIISVFLVCAQNF 120
OY 121 ROEYGAATGAVGAGVAGQLFGKSGSRVSMAGAVLGLGSKIGSQMDQDKIKLNQSL 180
DB 121 ROEYGAATGAVGAGVAGQLFGKSGSRVSMAGAVLGLGSKIGSQMDQDKIKLNQSL 180
OY 93 -----FGFASGRATL-----KKGQFTITQ----- 111
DB 93 -----FGFASGRATL-----KKGQFTITQ----- 111
OY 181 EYKAKAGVTR--WRNPDTGNSYSVEPRYTORYNKORRROOYCREFOKAMIAQOKEIY 238
DB 112 ----GRITKSDWSNNTOTNDYSFDASSSTPVVNP-----KVTGYIGAK--VL 153
OY 239 GTACPOPD 246
DB 154 GTA-PCPD 160

```

RESULT 8
AAR63634

ID AAR63634 standard; protein; 162 AA.

AC AAR63634;

DT 05-JUN-1995 (first entry)

DE Cellulose binding domain.

XX Cellulose binding domain; CBD.

OS Clostridium cellulovorans.

XX W09424158-A.

PD 27-OCT-1994.

XX 14-APR-1994; 94MO-US04132.

XX 14-APR-1993; 93US-0048164.

PA (REGC) UNIV CALIFORNIA.

XX (YISS) YISSUM RES & DEV CO.

XX Doi RH, Goldstein MA, Shoseyov O, Shplegl I;

XX WPI: 1994-34167/42.
 DR N-PSDB: AA072917.
 XX Isolated cellulose binding domain and fusion proteins - with
 PT applications, including drug delivery, affinity separations, and
 PT diagnostic techniques
 XX Claim 1, Fig 1, 125pp; English.
 PS
 XX AA072917/R63634 is a novel isolated cellulose binding domain. It pref.
 CC binds crystalline with a Kd ranging from 1.5 microm - 0.5 microm,
 CC pref. with a Kd of <1.2 microm, esp. <1.0 microm. The protein and
 CC the encoding DNA are claimed, as is a CBD fusion protein (FP)
 CC comprising the CBD and a second protein. The second protein is pref.
 CC Protein A, heparinase, a hormone or an enzyme capable of degrading
 CC an environmental pollutant, or an HSP, HSP antibody, cross-reactive
 CC HSP-related protein or an antigenic portion of this. The CBD and FP
 CC may be used in drug delivery, affinity seps. and diagnostic
 CC techniques. CBD nucleic acid may be obtd. from a variety of cell
 CC sources that produce CBDs that bind with high affinity and in a
 CC reversible manner or that produce CBD encoding mRNA. The preferred
 CC source of CBD encoding nucleic acid is C. cellulovorans.
 CC
 SQ Sequence 162 AA:
 Query Match 36.2%; Score 483; DB 15; Length 162;
 Best Local Similarity 51.7%; Pred. No. 2.7e-39; Mismatches 14; Indels 76; Gaps 4;
 Matches 106; Conservative 9;
 QY 1 MSVEFNSKNSAQTNSITPIIKITNTSDSLNDLVKRYRYTSDGTGQGFPCDHAGAL 60
 Db 6 MSVEFNSKNSAQTNSITPIIKITNTSDSLNDLVKRYRYTSDGTGQGFPCDHAGAL 65
 QY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYLDP SHMRGCLQSSLIISVFLVCAQNF 120
 Db 66 LGNSYVDNTSKVTANFVKETASPTSTYDYLDP SHMRGCLQSSLIISVFLVCAQNF 97
 QY 121 ROEYGAATGAVGVAGVQGFKGSGRVSMAGAVLGLGSKIGQSMDDQDKIKLNSL 180
 Db 98 -----FGFASGRATL-----KKGQFTITQ----- 116
 QY 181 EKVAGQVTR--WRNPDTGNSYSVE 203
 Db 117 -----GRITKSDMSNYQTNDYSFD 136
 RESULT 9
 AA05745
 ID AA05745 standard; Protein: 163 AA.
 XX
 AC AA05745;
 XX
 DT 24-SEP-2001 (first entry)
 XX
 DE Clostridium cellulovorans cellulose binding domain (CBDClos).
 XX
 KW Polysaccharide modification; polysaccharide binding domain; PBD; paper;
 KW yarn; fiber; textile; biological crosslinker; mechanical property;
 KW Clostridium cellulovorans cellulose binding domain; wet strength;
 KW durability; elasticity; CBDClos; cellulose binding protein A; CBP A.
 XX
 OS Clostridium cellulovorans.
 XX
 PN WO200134091-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 02-NOV-2000; 2000WO-1L00708.
 XX
 PR 08-NOV-1999; 99US-0164140.
 PR 18-NOV-1999; 99US-0166389.
 XX

PA (CBDT-) CBD TECHNOLOGIES LTD.
 PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
 XX
 PI Levy I, Nussimovitch A, Shoseyov O;
 XX
 DR WPI: 2001-457121/49.
 DR N-PSDB: AA011042.
 XX
 PT Preparation of a polysaccharide containing material having at least one
 PT desired structural, chemical, physical, electrical and/or mechanical
 PT property -
 XX
 PS Example 1.1: Page 111-112; 121pp; English.
 XX
 XX The present invention relates to methods and compositions for cross-
 CC linking and/or modifying the properties of polysaccharide materials.
 CC The method involves treating the polysaccharide structure with a
 CC polysaccharide binding domain (PBD) fusion protein. The method is
 CC used to alter the structural, chemical, physical, electrical and
 CC mechanical properties of polysaccharide materials such as paper,
 CC yarns, fibers and textiles, using biological crosslinking agents.
 CC The polysaccharide containing materials have improved mechanical
 CC properties such as wet strengths, durability and elasticity. The PBD
 CC reagent is applied in the forming stage in fusing paper manufacture
 CC which eliminates the recycling step. The use of a biological crosslinker
 CC improves the recyclability of paper products. The PBD reagent maintains
 CC the fine fibers in a slurry therefore resulting in better recovery of
 CC raw materials. The PBD molecules are eluted by strong alkaline conditions
 CC which enhances the ability of the alkaline glue used in binding to
 CC penetrate paper.
 CC The present sequence is Clostridium cellulovorans cellulose binding
 CC domain (CBDClos) of cellulose binding protein A (CBP A).
 CC
 SQ Sequence 163 AA:
 Query Match 36.2%; Score 483; DB 22; Length 163;
 Best Local Similarity 51.7%; Pred. No. 2.7e-39; Mismatches 14; Indels 76; Gaps 4;
 Matches 106; Conservative 9;
 QY 1 MSVEFNSKNSAQTNSITPIIKITNTSDSLNDLVKRYRYTSDGTGQGFPCDHAGAL 60
 Db 7 MSVEFNSKNSAQTNSITPIIKITNTSDSLNDLVKRYRYTSDGTGQGFPCDHAGAL 66
 QY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYLDP SHMRGCLQSSLIISVFLVCAQNF 120
 Db 67 LGNSYVDNTSKVTANFVKETASPTSTYDYLDP SHMRGCLQSSLIISVFLVCAQNF 98
 QY 121 ROEYGAATGAVGVAGVQGFKGSGRVSMAGAVLGLGSKIGQSMDDQDKIKLNSL 180
 Db 99 -----FGFASGRATL-----KKGQFTITQ----- 117
 QY 181 EKVAGQVTR--WRNPDTGNSYSVE 203
 Db 118 -----GRITKSDMSNYQTNDYSFD 137
 RESULT 10
 AA05748
 ID AA05748 standard; Protein: 428 AA.
 XX
 AC AA05748;
 XX
 DT 24-SEP-2001 (first entry)
 XX
 DE Clostridium cellulovorans protein A-cellulose binding domain.
 XX
 KW Polysaccharide modification; polysaccharide binding domain; PBD; paper;
 KW yarn; fiber; textile; biological crosslinker; mechanical property;
 KW wet strength; durability; elasticity; protein A-cellulose binding domain;
 KW ProDA-CBD.
 XX
 OS Clostridium cellulovorans.
 XX

QY	121	ROEVGATTAAGVAGAGQLPGKSGSRVNAIGCAVLGGLIGRIGSGMDQDQIKLNQSL	180
DB	362	-----FFSGFAGRATL-----KKGFITITD-----	380
QY	181	EKKVAGQVTR--WRNPDIGNSYSVE	203
DB	381	-----GRITKSDMSNTYTQINDYSED	400
RESULT 11			
ID	AAV39952	standard. Protein; 382 AA.	
XX	AAV39952;		
XX	AC		
XX	DT	13-DEC-1999 (first entry)	
XX	XX	Gaussia luciferase fusion protein sequence.	
DE	XX	Luciferase; green fluorescent protein; GFP; screening assay; diagnosis;	
XX	XX	bioluminescence-generating system; toy; cosmetic; fairy dust; beverage;	
KW	XX	body paint; squirt gun; balloon; slimy play material; soap; toothpaste;	
KM	XX	fusion protein.	
KW	XX	Gaussia sp.	
OS	XX	WO9949019-A2.	
PN	XX	30-SEP-1999.	
XX	XX	26-MAR-1999; 99WO-US06698.	
PF	XX	27-MAR-1998; 98US-0076624.	
PR	XX	15-JUN-1998; 98US-0089367.	
PR	XX	01-OCT-1998; 98US-0102339.	
PR	XX	(PROL-) PROLUME LTD.	
PA	XX	(BRYA/) BRYAN B J.	
PA	XX	Bryan BJ, Szent-Gyorgyi C;	
PI	XX	WPI: 1999-580443/49.	
XX	XX	N-PSDB; AA227550.	
DR	XX	New isolated Renilla mulleri, Gaussia and Pleuromamma luciferase and	
PT	XX	Renilla and Ptilosarcus green fluorescent protein nucleic acids -	
XX	XX	Disclosure: Page 222-223; 23pp; English.	
PS	XX		
XX	XX	This sequence represents a luciferase of the invention. The invention	
XX	XX	relates to Renilla mulleri, Gaussia and Pleuromamma luciferase and	
CC	XX	Renilla and Ptilosarcus green fluorescent protein (GFP) nucleic acids and	
CC	XX	proteins. The luciferases and GFPs can be used in	
CC	XX	bioluminescence-generating systems, assays, screening methods, diagnostic	
CC	XX	method and articles of manufacture. They can be expressed using	
CC	XX	e.g. bacterial, yeast, fungal, plant, insect or animal cells. The Renilla	
CC	XX	mulleri, Gaussia and, Pleuromamma luciferase or Renilla or Ptilosarcus	
CC	XX	GFPs can be used in e.g. toys, cosmetics, fountains, personal care items,	
CC	XX	fairy dust, beverages, soft drinks, foods, textile products, bubbles,	
CC	XX	balloons, personal items, dentifrices, soaps, body paints, bubble bath,	
CC	XX	ink or paper products. In particular, they can be used in e.g. squirt	
CC	XX	guns, pellet guns, finger paints, foot bags, greeting cards, slimy play	
CC	XX	material, clothing, bubble making toys, bath powders, cosmetics, body	
CC	XX	lotions, gels, body powders, body creams, toothpastes, mouthwashes,	
CC	XX	soaps, body paints, bubble bath, inks, wrapping paper, gelatins, icings,	
CC	XX	frostings, greeting cards, beer, wine, champagne, soft drinks, ice cubes,	
CC	XX	ice, dry ice or fountains. The nucleic acids can also be used to produce	
CC	XX	transgenic fish and plants.	
SO	Sequence	382 AA;	
Query Match	36.0%;	Score 480.5;	DB 20; Length 382;
Best Local Similarity	62.6%;	Pred. No. 1.6e-38;	

Matches 107; Conservative 10; Mismatches 33; Indels 21; Gaps 4;

QY 1 MSVEFYNSKSAQTNSITPIIKITNTSDSLNLDVKKVRYTSDGTGOTFWCHAGAL 60
 DB 1 MSVEFYNSKSAQTNSITPIIKITNTSDSLNLDVKKVRYTSDGTGOTFWCHAGAL 60
 QY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYDYE : : : :
 DB 61 LGNSYVDNTSKVTANFVKETASPTSTYDYDYEFCFASGATLKKQCFITIOGRITKSDMS 120
 QY 118 NESRO-----EVGAATGAVGVAGOLFGKSGRVSATGCA-VLGGILGS 162
 DB 121 NTGTNDISFDASSSTPYVNPKYTG-----ISGAKVLTARGS 159

RESULT 12

AAM90077 standard; Protein: 162 AA.

AC AAM90077;

DT 09-MAR-1999 (first entry)

DE C. cellulovorans Cbpa CBD protein.

XX Cellulose binding domain; CBD; Cbpa; fusion protein; diagnosis; HSP;
 KW immunosassay; heat-shock protein; cross reactive protein; detection;
 KM antigenic fragment; antibody; insulin-dependent diabetes mellitus;
 KM cellulose; chitin; cellulytic; amorphogenic.

XX Clostridium cellulovorans.

PN US5856201-A.

XX 05-JAN-1999.

XX 27-OCT-1994; 94US-0330394.

XX 27-OCT-1994; 94US-0330394.

XX 14-APR-1993; 93US-0048164.

XX 14-APR-1994; 94WO-US04132.

XX (REGC) UNIV CALIFORNIA.

XX (YISS) YISSUM RES & DEV CO.

XX Doi RH, Goldstein MA, Shoseyov O, Shplegl I;

XX WPI, 1999-105130/09.

XX N-PSDB: AAV74072.

XX detection of a specific analyte by reaction with binding agent fused
 PT to cellulose binding domain - and subsequent treatment with
 PT cellulose and reaction of insoluble product with a label specific
 PT for the analyte

XX Claim 3; Fig 1A-B; 63pp; English.

XX This sequence represents a cellulose binding domain (CBD) derived from
 CC the Cbpa protein of Clostridium cellulovorans. The sequence is used the
 CC construction of a fusion protein which can be used in diagnostic
 CC immunosassays, e.g. to detect heat-shock proteins (HSP) and their
 CC cross-reactive proteins, antigenic fragments or HSP-specific antibodies
 CC (which indicate insulin-dependent diabetes mellitus, or susceptibility
 CC to it. The CBD, has very high affinity for cellulose (including
 CC crystalline forms) and chitin (dissociation constant 0.8-1.4 mu M), but
 CC has almost no cellulytic or amorphogenic activities. The CBD binds over
 CC a wide pH range and is not released from cellulose by washing with water.

XX Sequence 162 AA;

Query Match 35.7%; Score 477; DB 20; Length 162;
 Best Local Similarity 97.8%; Pred. No. 1e-38;
 Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVEFYNSKSAQTNSITPIIKITNTSDSLNLDVKKVRYTSDGTGOTFWCHAGAL 60
 DB 6 MSVEFYNSKSAQTNSITPIIKITNTSDSLNLDVKKVRYTSDGTGOTFWCHAGAL 65

QY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYDYE 92
 DB 66 LGNSYVDNTSKVTANFVKETASPTSTYDYDYE 97

RESULT 13

AAM90081 standard; Protein: 154 AA.

AC AAM90081;

DT 09-MAR-1999 (first entry)

DE C. cellulovorans CBD-Kpni fusion protein.

XX Cellulose binding domain; CBD; Cbpa; fusion protein; diagnosis; HSP;
 KW immunosassay; heat-shock protein; cross reactive protein; detection;
 KM antigenic fragment; antibody; insulin-dependent diabetes mellitus;
 KM cellulose; chitin; cellulytic; amorphogenic.

XX Clostridium cellulovorans.

XX Synthetic.

PN US5856201-A.

XX 05-JAN-1999.

XX 27-OCT-1994; 94US-0330394.

XX 27-OCT-1994; 94US-0330394.

XX 14-APR-1993; 93US-0048164.

XX 14-APR-1994; 94WO-US04132.

XX (REGC) UNIV CALIFORNIA.

XX (YISS) YISSUM RES & DEV CO.

XX Doi RH, Goldstein MA, Shoseyov O, Shplegl I;

XX WPI, 1999-105130/09.

XX detection of a specific analyte by reaction with binding agent fused
 PT to cellulose binding domain - and subsequent treatment with
 PT cellulose and reaction of insoluble product with a label specific
 PT for the analyte

XX Claim 5; Column 55-56; 63pp; English.

XX This sequence represents a fusion protein (CBD-Kpni) composed of the Cbpa
 CC protein cellulose binding domain (CBD) from Clostridium cellulovorans
 CC This protein is used in diagnostic immunosassays, e.g. to detect
 CC heat-shock proteins (HSP) and their cross-reactive proteins, antigenic
 CC fragments or HSP-specific antibodies (which indicate insulin-dependent
 CC diabetes mellitus, or susceptibility to it. The CBD has very high
 CC affinity for cellulose (including crystalline forms) and chitin
 CC (dissociation constant 0.8-1.4 mu M), but has almost no cellulytic or
 CC amorphogenic activities. The CBD binds over a wide pH range and is not
 CC released from cellulose by washing with water.

XX Sequence 154 AA;

Query Match 35.4%; Score 472; DB 20; Length 154;
 Best Local Similarity 97.8%; Pred. No. 3e-38;
 Matches 89; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 SVEFYNSKSAQTNSITPIIKITNTSDSLNLDVKKVRYTSDGTGOTFWCHAGAL 61
 DB 1 SVEFYNSKSAQTNSITPIIKITNTSDSLNLDVKKVRYTSDGTGOTFWCHAGAL 60

OY 62 GNSYVDNTSKVTANFYKETASPTSTYDTYLD 92
 DB 61 GNSYVDNTSKVTANFYKETASPTSTYDTYVE 91

RESULT 14

AAW90080
 ID AAW90080 standard; Protein: 156 AA.

XX AAW90080;

AC 09-MAR-1999 (first entry)

DE C. cellulovorans CBD-HSP fusion protein.

KM Cellulose binding domain; CBD; Cbpa; fusion protein; diagnosis; HSP;

KW Immunassay; heat-shock protein; cross reactive protein; detection;

KW antigenic fragment; antibody; insulin-dependent diabetes mellitus;

XX cellulose; chitin; cellulolytic; amorphogenic.

OS Clostridium cellulovorans.

XX Synthetic.

PN US5856201-A.

PD 05-JAN-1999.

XX 27-OCT-1994; 94US-0330394.

XX 27-OCT-1994; 94US-0330394.

PR 14-APR-1993; 93US-0048164.

PR 14-APR-1994; 94WO-US04132.

XX (RECC) UNITV CALIFORNIA.

PA (YISS) YISSUM RES & DEV CO.

PI Dol RH, Goldstein MA, Shoseyov O, Shplegl I;

XX WPI: 1999-105130/09.

DR WPI: 1999-105130/09.

XX WPI: 1999-105130/09.

PT Detection of a specific analyte by reaction with binding agent fused

PT to cellulose binding domain - and subsequent treatment with

PT cellulose and reaction of insoluble product with a label specific

PT for the analyte

XX Claim 4: Column 51-52: 63pp; English.

PS This sequence represents a fusion protein composed of the Cbpa protein

XX cellulose binding domain (CBD) from Clostridium cellulovorans and a heat

CC shock protein (HSP) fragment. This protein is used in diagnostic

CC immunassays, e.g. to detect heat-shock proteins (HSP) and their

CC cross-reactive proteins, antigenic fragments or HSP-specific antibodies

CC to which indicate insulin-dependent diabetes mellitus, or susceptibility

CC to it. The CBD has very high affinity for cellulose (including

CC crystalline forms) and chitin (dissociation constant 0.8-1.4 mu M), but

CC has almost no cellulolytic or amorphogenic activities. The CBD binds over

CC a wide pH range and is not released from cellulose by washing with water.

RESULT 15
 AAR95080
 ID AAR95080 standard; peptide: 167 AA.

XX AAR95080;

AC 30-OCT-1996 (first entry)

DE Cellulose binding domain.

KM Cellulose binding domain; CBD; hapten; moiety; biotin; avidin;

KW streptavidin; affinity chromatography; cell separation,

KW cell immobilisation; protein immobilisation; enzyme immobilisation;

KW multienzyme reactors; signal immunoassays; drug delivery; pesticide;

XX cellulose; chitin.

OS Clostridium thermoceilum.

XX Key Location/Qualifiers

FT Misc-difference 139 /note= "Unidentified amino acid."

XX WO9613524-A1.

PD 09-MAY-1996.

XX 26-OCT-1995; 95WO-US13813.

XX 27-OCT-1994; 94IL-0111415.

XX (TECR) TECHNITION RES & DEV FOUND LTD.

PA (UYRA-) UNITV RAMOT APPLIED RES & IND DEV LTD.

PA (YEDA) YEDA RES & DEV CO LTD.

XX Bayer EA, Lamed R, Morag E, Shoham Y, Wilchek M;

PI WPI: 1996-239453/24.

DR WPI: 1996-239453/24.

XX WPI: 1996-239453/24.

PT Modified cellulose-binding domain moles - having attached hapten,

PT partic. biotin, to provide an affinity system for, e.g. sepsis,

XX assays, reactors, delivery etc.

XX Claim 12: Page 34-35: 53pp; English.

PS A modified cellulose binding domain (CBD) or fraction of it, linked

XX to a hapten moiety via one or more Cys or Lys residues and complexes

CC comprising the biotinylated CBD and a biotin-binding molecule

CC selected from modified or unmodified avidin or streptavidin or an

CC anti-biotin antibody, can be used in affinity chromatography, cell

CC separation, cell, protein and enzyme immobilisation, selective

CC removal of biological materials, multienzyme reactors, signal

CC immunoassays and drug (e.g. pesticide) delivery. The CBD molecule

CC can be modified with the hapten without affecting its high affinity

XX for cellulose, or chitin.

SO Sequence 167 AA;

Query Match 18.4%; Score 245; DB 17; Length 167;

Best Local Similarity 51.5%; Pred. No. 6e-16;

Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;

OY 1 MSVEFYNSKSAQTNSITPIIKITNTSDSLNDVKVRYRYYTSDGOTGTFMCDHAGAL 60

DB 9 LKVEFYNSKSAQTNSITPIIKITNTSDSLNDVKVRYRYYTSDGOTGTFMCDHAGAL-AI 67

OY 61 LGN--SYVDNTSKVTANFYKETASPTSTYDTYLDPSHMRCLQ 101

DB 68 ICSNGSYNGTITSNVKGTFVKMSSS-TNNADTYLEISTGTGLE 109

Search completed: October 27, 2002, 10:57:58

Job time : 36.487 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 11:00:01 : Search time 28.7392 Seconds
(without alignments)
1540.986 Million cell updates/sec

Title: US-09-677-374-6
Perfect score: 1335
Sequence: 1 MSVEFYNSKSAQTNSITPL.....ITGTACPPDGRMQVISTEK 256

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriaph:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	815	61.0	162	2	Q9F9K8
2	303.5	22.7	148	2	054381
3	283	21.2	159	2	09F9P2
4	258.5	19.4	137	2	052252
5	252.5	18.9	144	2	09K2M6
6	251.5	18.8	137	2	031065
7	251.5	18.8	144	2	09K4W8
8	249	18.7	154	2	045996
9	244	18.3	154	2	053154
10	239	17.9	151	2	09F9Q9
11	236.5	17.7	131	2	09F001
12	236.5	17.7	131	2	052657
13	235.5	17.6	131	2	09L532
14	233.5	17.5	1162	2	082830
15	223.5	16.7	1483	16	0977Y4
16	216.5	16.2	105	2	031208

17	195.5	14.6	1915	2	Q9RPL0	Q9RPL0 acetivibrio
18	193	14.5	887	2	Q9L3J8	Q9L3J8 clostridium
19	169.5	12.7	321	2	Q46392	Q46392 clostridium
20	152.5	11.4	997	2	Q924I1	Q924I1 bacillus sp
21	151.5	11.3	1779	2	052374	052374 caldicellul
22	150.5	11.3	1770	2	09X3P5	09X3P5 caldicellul
23	149.5	11.2	261	2	09AG07	09AG07 caldicellul
24	149.5	11.2	1426	2	09X3P6	09X3P6 caldicellul
25	148.5	11.1	1751	2	09AG04	09AG04 caldicellul
26	144.5	10.8	996	2	Q9AQH0	Q9AQH0 caldicellul
27	142.5	10.7	1000	2	024820	024820 thermophil
28	142	10.6	1711	2	P96311	P96311 anaerocellu
29	139	10.4	77	2	Q9AC07	Q9AC07 rickettsia
30	138.5	10.4	921	2	Q9A8L8	Q9A8L8 caldibacill
31	137	10.3	199	16	Q985G4	Q985G4 rhizobium
32	135.5	10.1	930	2	Q9RFX5	Q9RFX5 caldibacill
33	128	9.6	170	2	Q9RFX6	Q9RFX6 caldibacill
34	127.5	9.6	182	16	Q9HX13	Q9HX13 pseudomonas
35	122	9.1	257	16	Q9A8M8	Q9A8M8 caulobacter
36	121	9.1	499	2	Q93RT6	Q93RT6 bacillus su
37	121	9.1	508	2	Q93LD0	Q93LD0 bacillus su
38	120.5	9.0	136	16	Q92R89	Q92R89 rhizobium m
39	114.5	8.6	501	2	083012	083012 bacillus sp
40	114	8.5	499	2	052731	052731 bacillus sp
41	113.5	8.5	499	2	045532	045532 bacillus su
42	111.5	8.4	154	16	Q914S1	Q914S1 pseudomonas
43	111.5	8.4	2316	2	Q9F0J9	Q9F0J9 bacteroides
44	111	8.3	155	2	Q9F6B1	Q9F6B1 edwardsiell
45	110	8.2	155	2	Q9RB08	Q9RB08 pectobacter

ALIGNMENTS

RESULT 1

Q9F9K8

PRELIMINARY: PRT: 162 AA.

AC Q9F9K8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE 17 KDA ANTIGEN.
GN OSPA.
OS Piscirickettsia salmonis.
OC Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;
OC Piscirickettsia.
OX NCBI_TaxID=1238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LF-89;
RA Kuzuk W.A., Burton J., Thornton J.C., Kay W.W.;
RT "Identification of a genus-common Rickettsial surface antigen in the
RT salmonid pathogen Piscirickettsia salmonis.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF184152; AAC17000.1; -;
SQ SEQUENCE 162 AA; 17661 MW; DDE99E6FD94A527E CRC64;

Query Match 61.0%; Score 815; DB 2; Length 162;

Best Local Similarity 98.8%; Pred. No. 3.6e-59;
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 97 RGCAGSSLLIISVFLVGCACNFRSROEVGATGAVGVAGQALFGKSGSRVMAIGAVL 156
DB 3 RGLDGGSSLLIISVFLVGCACNFRSROEVGATGAVGVAGQALFGKSGSRVMAIGAVL 62
QY 157 GGLGSKRIGSGMDQDDKIKLNSLEKVKAGOVTRMRPDTGNSYSVPRTYQRYNKOER 216
DB 63 GGLGSKRIGSGMDQDDKIKLNSLEKVKAGOVTRMRPDTGNSYSVPRTYQRYNKOER 122
QY 217 RQYVCREFOQKAMTAGQKQETIYGTACPODGRMOWISTEK 256
DB 123 RQYVCREFOQKAMTAGQKQETIYGTACPODGRMOWISTEK 162

RESULT 2

054381 ID PRELIMINARY; PRT: 148 AA.

AC 054381; 01-JUN-1998 (TREMblrel. 06, Created)

DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)

DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)

DE 17 KDA COMMON-ANTIGEN (FRAGMENT).

OS Rickettsia sp.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=789;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98087556; PubMed=9425244;

RA Davis M.J., Ying Z., Brunner B.R., Pantoja A., Ferwerda F.H.;

RT "Rickettsial relative associated with papaya bunchy top disease.";

RL Curr. Microbiol. 36:80-84(1998).

DR EMBL; 076907; AAC02809.1; -.

FT NON_TER 1 148

SQ SEQUENCE 148 AA; 15050 MW; A7AFEEFE0AE6E4C CRC64;

Query Match 22.7%; Score 303.5; DB 2; Length 148;

Best local similarity 40.1%; Pred. No. 2.2e-17;

Matches 55; Conservative 30; Mismatches 43; Indels 9; Gaps 2;

QY 120 SROEVGATGAVVGVAGOLFEGKSGRVSMAIGAVLGLIGSKRTGSGMDQDK---IK 175.

DB 17 NKGSGTIGLIGLIGVSGFGGTGLAAGALGALILGAGMDEQDKRLALT 76

QY 176 LMSLEKVRAGQVTRNRPDGTNSVPEVRYRYNKOEROQYCRFEOQKAMTAGQK 235

DB 77 SQRLEAPSGSSVYMNPNNGNYGVTPSKAV-----KNNTGYCRREYQTVVVGKQ 131

QY 236 EHYGTACPOPDGRWOV 252

DB 132 KAYGTACROPDGMQOV 148

RESULT 3

09F9F2 ID PRELIMINARY; PRT: 159 AA.

AC 09F9F2; 01-MAR-2001 (TREMblrel. 16, Created)

DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)

DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)

DE 17 KDA GENUS-COMMON ANTIGEN.

OS Rickettsia felis (Rickettsia azadi).

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=42862;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21217364; PubMed=11321078;

RA Bouyer D.H., Stenos J., Crocquet-Valdes P., Moron C.G., Popov V.L.,

RA Zavala-Velazquez J.E., Foil L.D., Stothard D.R., Azad A.F.,

RA Walker D.H.;

RT "Rickettsia felis: molecular characterization of a new member of the

RT spotted fever group.";

RL Int. J. Syst. Evol. Microbiol. 51:339-347(2001).

DR EMBL; A1195118; AAG28452.1; -.

SQ SEQUENCE 159 AA; 16457 MW; 34C5B020AF470A1F CRC64;

Query Match 21.2%; Score 283; DB 2; Length 159;

Best local similarity 37.9%; Pred. No. 1.1e-15;

Matches 61; Conservative 30; Mismatches 34; Indels 16; Gaps 6;

QY 103 SLLIISV---FLVGC--AQNFRGVGATGAVVGVAGOLFEGKSGRVSMAIG-GAVL 156

DB 5 SKMIITLAAAMLQACNGPGMKNQGTGTLGAGCALLGSGFGKKGQL-VGVGVGALL 63

QY 157 GGLIGSKIGSMQDQK---IKLMSLEKVRAGQVTRNRPDGTNSVPEVRYRYN 212

DB 64 GAVLIGGIGAGMDEQDKRLAELTSGRLEATPDSGTSVEKRNPNHGVYVNPNTY----- 119

QY 213 KQERROQYCRFEOQKAMTAGQKQKQYGTACPOPDGRWOV 253

DB 120 -RNSTGYCRREYQTVVVGKQKAYGNACROPDGMQOVN 159

RESULT 4

052252 ID PRELIMINARY; PRT: 137 AA.

AC 052252; 01-JUN-1998 (TREMblrel. 06, Created)

DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)

DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)

DE 17 KDA ANTIGEN (FRAGMENT).

OS Rickettsia cooley.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=69410;

RN [1]

RP SEQUENCE FROM N.A.

RA Billings A.N., Telow G.J., Walker D.H.;

RT "Molecular characterization of a novel spotted fever group rickettsial

RT species from ixodes scapularis in Texas.";

RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.

DR EMBL; AF031534; AAB95267.1; -.

FT NON_TER 1 14215 MW; A27597A9FDB5FC3 CRC64;

SQ SEQUENCE 137 AA; 14215 MW; A27597A9FDB5FC3 CRC64;

Query Match 19.4%; Score 258.5; DB 2; Length 137;

Best local similarity 39.3%; Pred. No. 9.4e-14;

Matches 53; Conservative 25; Mismatches 46; Indels 11; Gaps 4;

QY 117 QNFSROEVGATGAVVGVAGOLFEGKSGRVSMAIG-GAVLGLIGSKRTGSGMDQDK-- 173

DB 7 KGMKQGTGTLGAGCALLGSGFGKQQL-VGVGVGALLGAVLIGQIGAMDEQDKRL 65

QY 174 --IKLMSLEKVRAGQVTRNRPDGTNSVPEVRYRYNKOEROQYCRFEOQKAMTA 231

DB 66 AELTSGRLEATPDSGTSVEKRNPNHGVYVNPNTY-----RNSTGYCRREYQTVVIG 120

QY 232 GQKQKAYGNACROPD 246

DB 121 GKQKAYGNACROPD 135

RESULT 5

09K2N6 ID PRELIMINARY; PRT: 144 AA.

AC 09K2N6; 01-OCT-2000 (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)

DE 17KDA ANTIGEN (17 KDA ANTIGEN) (FRAGMENT).

OS male-killing Rickettsia from Adalia bipunctata.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=38028;

RN [1]

RP SEQUENCE FROM N.A.

RA Schlenker H.J.G.V.D., Habig M., Stoggett J.J., Webberley M.K.,

RA Bertrand D., Hurst G.D.D., Majerus M.E.N.;

RT "On the evolution of male-killing: Monophyletic origin and horizontal

RT transfer of male-killing Rickettsia (a-Proteobacteria) from two con-

RT generic ladybirds, Adalia bipunctata L. and A. decempunctata L.

RT (Coleoptera: Coccinellidae).";

RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL; AJ269518; CAB96383.1; -.

DR EMBL; AJ269517; CAB96382.1; -.

FT NON_TER 1 144

[illegible]

```
DE    17 KDA ANTIGEN (FRAGMENT).  
OS    male-killing Rickettsia from Adalia decempunctata.  
OC    Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC    Rickettsiaceae; Rickettsiae; Rickettsia.  
OX    NCBI_TaxID=120393;  
RN    [1]  
RP    SEQUENCE FROM N.A.  
RX    MEDLINE=20575219; PubMed=11133455;  
RA    Schultenbury H.J.G.V.D., Habig M., Sjoggett J.J., Webberley M.K.,  
RA    Bertrand D., Hurst G.D., Majerus M.E.N.;  
RT    "Incidence of male-killing Rickettsia spp. (alpha-Proteobacteria) In  
RT    the Ten-spot Ladybird Beetle Adalia decempunctata L. (coleoptera:  
RT    Coccinellidae)." ;  
RL    Appl. Environ. Microbiol. 67:270-277(2001).  
DR    EMBL; AJ269516; CAB96381.1; -.  
FT    NON_TER      1         144       1  
SO    SEQUENCE     144 AA;   14801 MW;   CB25472F16A56AE7 CRC64;  
  
Query Match                               18.8%; Score 251.5; DB 2; Length 144;  
Best Local Similarity                      39.4%; Pred. No. 3,7e-13;  
Matches 52; Conservative % 24; Mismatches 45; Indels 11; Caps  
  
QY    120 SRQEGATGCAVGCAGQLRKSGSRSMALIG-GAVLGILSKIQSQMPODK-----I 170  
DB    17 NKGGTGLLGGAGALLGSQGKRGDL-VGVGVALGAVALGGDAGMDOPRRAL 75  
OY    175 KLNSLEKVKKGQGYTRWNPPTGNSYSVEPRTRYNKKORRQCYCREFOOKMIAGOK 234  
DB    76 TSQRLEAPSPSNGVENRNPNGNHGYTPPKTY-----RNSTQCYREHYQTIVYGKQ 130  
QY    235 QEITYGTACQPDP 246  
DB    131 QKSYGNACQPD 142  
  
RESULT 8  
Q45996 PRELIMINARY; PRT; 1546 AA.  
AC    O45996;  
DT    01-NOV-1996 (TREMBLrel_01, Created)  
DT    01-MAR-2001 (TREMBLrel_16, Last sequence update)  
DT    01-JUN-2001 (TREMBLrel_17, Last annotation update)  
DE    SCAFOLDING PROTEIN PRECURSOR.  
GN    ClpC.  
OS    Clostridium cellulolyticum.  
OC    Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC    Clostridii;  
OX    NCBI_TaxID=1521;  
RN    [1]  
RP    SEQUENCE FROM N.A.  
RC    STRAIN=ATCC 35319;  
RX    MEDLINE=96218696; PubMed=8636029;  
RA    Pages S., Belaich A., Tardif C., Reverbel-Jeroy C., Gaudin C.,  
RA    Belaich J.P.;  
RT    "Interaction between the endoglucanase CelA and the scaffolding  
RT    protein ClpC of the Clostridium cellulolyticum celulosome." ;  
RT    J. Bacteriol. 178:2279-2286(1996).  
RN    [2]  
RP    SEQUENCE FROM N.A.  
RC    STRAIN=ATCC 35319;  
RX    MEDLINE=99173902; PubMed=10074072;  
RA    Pages S., Belaich A., Fierobe H.P., Tardif C., Gaudin C.,  
RA    Belaich J.P.;  
RT    "Sequence analysis of scaffolding protein ClpC and ORFXp, a new  
RT    cohesin-containing protein in Clostridium cellulolyticum: comparison  
RT    of various cohesin domains and subcellular localization of ORFXp." ;  
RL    J. Bacteriol. 181:1801-1810(1999).  
RN    [3]  
RP    SEQUENCE FROM N.A.  
RC    STRAIN=ATCC 35319;  
RA    Reverbel-Jeroy C., Tardif C., Belaich A., Bernadac A., Gaudin C.,  
RA    Belaich J.
```

"Molecular study and overexpression of the Clostridium cellulolyticum celC gene in *Escherichia coli*.".
 RT Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35319;
 RA Pages S., Belaich A., Reverbel C., Tardif C., Fierobe H.P., Gaudin C., Belaich J.P.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RT EMBL: U40345; AAC28899.2; -
 DR HSSP: Q06851; INBC.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR002102; Cohesin.
 DR InterPro: IPR003880; Phosphopant_attach.
 DR Pfam: PF00942; CBD_3; 1.
 DR Pfam: PF00963; Cohesin; 8.
 DR Prodom: PD001947; CBD_3; 1.
 DR PROSITE: PS00012; PHOSPHOPANTHEINE; UNKNOWN_1.
 DR Signal.
 RT SIGNAL.
 FT CHAIN
 SQ SEQUENCE 1546 AA; 158748 MW; F8651504EC27809F CRC64;

Query Match 18.7%; Score 249; DB 2; Length 1546;
 Best local similarity 33.8%; Pred. No. 1.3e-11;
 Matches 72; Conservative 32; Mismatches 81; Indels 28; Gaps 7;

OY 1 MSVEFYNSKSKQNTSIPITIKITNTSDILNDVKYRYTSDGTGCGTFCWCDHAGAL 60
 DB 33 VSGVFNNKSSPASNSIYARFVNTSGPILADLKRIYTTODADKPLTFWCDHAGYM 92
 OY 61 LGNSYVDNTSKYTANFVKETASTYDYLD--PSHMRCLGSSLIITISVLYGCAQ 117
 DB 93 SGNNTIDATSKYGTGR-KAVSPAVFNADHYLEVALNSDAGSLPAGSGIEIQTFRANDWS 151
 OY 118 NFSROEVAATGAVVGVAGOLFGRKSGS---GRVSMATGGAVLGGLISKISQSDQDQK 173
 DB 152 NFDQSDNWSYTA-----GSYMDQKISAFVGTGLAYG-STPDGCPNPDP 197
 OY 174 IKLNOSLEKVKAG-----QVTRMRNPDGTGNSYS 201
 DB 198 -TINPTISAKAGSPADTKITLTPNGTNGIS 229

RESULT 9
 053154 PRELIMINARY; PRT; 154 AA.

AC 053154; (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1996 (TREMBlrel. 08, Last annotation update)
 DE (CLONE PRB FTSE 1), 5' END CDS (FRAGMENT).
 OS Rickettsia sp.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 NCBI_TaxID=789;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93084757; PubMed=1452660.
 RA Baird R.W., Lloyd M., Stenos J., Ross B.C., Stewart R.S., Dwyer B.;
 RT "Characterization and comparison of Australian human spotted fever group rickettsiae."
 RL J. Clin. Microbiol. 30:2896-2902(1992).
 DR EMBL: M93931; AAA73386.1; -
 FT NON_TER 154
 SQ SEQUENCE 154 AA; 15849 MW; F5C35855EDB439D2 CRC64;

Query Match 18.3%; Score 244; DB 2; Length 154;
 Best local similarity 35.3%; Pred. No. 1.7e-12;
 Matches 55; Conservative 31; Mismatches 54; Indels 16; Gaps 6;

DB 5 SKTMTIALATSLHQAACNGPGMKGOTGTLGAGAGALLGSGFGKGOQL-VGVGVGALL 63
 OY 157 GGLISKISQSDQDQK----IKLNOSLEKVKAGVTRMRNPDGTGNSYSVPEPRTYGRYN 212
 DB 64 GAVLGQIGAGMEDQDRRLAELTSGRALETAPSGSVNEWRPNDSYGVYTPNKTYRST 123
 OY 213 KQERRQYCREFOOKAMIAQKOEIYGTACPPDGR 248
 DB 124 GQD-----CRVYTYVYIGSKQKAGNACRPPDQ 154

RESULT 10
 09F909 PRELIMINARY; PRT; 151 AA.

AC 09F909; (TEMBLrel. 16, Created)
 DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
 DE OUTER MEMBRANE PROTEIN (FRAGMENT).
 OS Rickettsia helvetica.
 OC Bacteria; proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 NCBI_TaxID=35789;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nilsson K., Pahlson C.;
 RT "Novel peptide diagnostic reagent and kit for detection of Rickettsiosis."
 RT Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF181036; AAC09427.1; -
 FT NON_TER 151
 SQ SEQUENCE 151 AA; 15621 MW; B77407B9C71E4B39 CRC64;

Query Match 17.9%; Score 239; DB 2; Length 151;
 Best local similarity 35.9%; Pred. No. 4.2e-12;
 Matches 55; Conservative 28; Mismatches 54; Indels 16; Gaps 6;

OY 103 SLLIITSV---FLVGC--AQNFSROEVAATGAVVGVAGOLFGRKSGRVSMAIG-GAVL 156
 DB 5 SKTMTIALASMLQAACNGPGMKGOTGTLGAGAGALLGSGFGKGOQL-VGVGVGALL 63
 OY 157 GGLISKISQSDQDQK----IKLNOSLEKVKAGVTRMRNPDGTGNSYSVPEPRTYGRYN 212
 DB 64 GAVLGQIGAGMEDQDRRLAELTSGRALETAPSGSVNEWRPNDSYGVYTPNKTYRST 119
 OY 213 KQERRQYCREFOOKAMIAQKOEIYGTACPPDGR 245
 DB 120 -RNSTGQYCREYTYVYIGSKQKAGNACRPPDQ 151

RESULT 11
 09F001 PRELIMINARY; PRT; 131 AA.

AC 09F001; (TEMBLrel. 16, Created)
 DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
 DE 17 KDA PROTEIN (FRAGMENT).
 OS Rickettsia sp. California 2.
 OC Bacteria; proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 NCBI_TaxID=147259;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CALIFORNIA 2;
 RA Raoult D.;
 RT "A new SFG rickettsia isolated from fleas."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CALIFORNIA 2;
 RA Roux V., Raoult D.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 11:00:36 ; Search time 9.72712 Seconds
(without alignments)
1019.028 Million cell updates/sec

Title: US-09-677-374-6
Perfect score: 1335
Sequence: 1 MSVEFYNSKNSAQTNSITPI.....IYGTACPDGMRQVISTEK 256

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	487	36.5	1848	1	CBPA_CLOCL
2	288	21.6	139	1	17KD_RICPR
3	286	21.4	159	1	17KD_RICJA
4	285	21.3	159	1	17KD_RICCN
5	276.5	20.7	159	1	17KD_RICRY
6	268	20.1	154	1	17KD_RICAU
7	259	19.4	154	1	17KD_RICPA
8	259	19.4	154	1	17KD_RICRH
9	256	19.2	154	1	17KD_RICMO
10	255	19.1	154	1	17KD_RICAM
11	245	18.4	1853	1	CITPA_CLOTM
12	245	18.4	1853	1	CITPA_CLOTM
13	188.5	14.1	879	1	GUNT_CLODM
14	188	14.1	986	1	GUNT_CLOSR
15	162.5	12.2	914	1	GUX2_CLOSR
16	156	11.7	80	1	17KD_RICCA
17	143.5	10.7	1039	1	GUNB_CALSA
18	143.5	10.7	1331	1	MANB_CALSA
19	138.5	10.4	1742	1	GUNA_CALSA
20	121	9.1	499	1	GUN2_BACSU
21	115	8.6	499	1	GUN3_BACSU
22	112.5	8.4	155	1	PCP_YEREN
23	111	8.3	145	1	YCEA_PAEIA
24	109.5	8.2	499	1	GUNT_BACSU
25	109	8.2	153	1	SLYB_SALTY
26	104.5	7.8	155	1	SLYB_ECOLI
27	103	7.7	504	1	GUNW_ERMCA
28	101.5	7.6	155	1	PCP_HAEIN
29	101.5	7.6	505	1	GUNV_ERMCA
30	100.5	7.5	700	1	GUNA_PAEIA
31	100	7.5	179	1	YCEJ_ECOLI
32	100	7.5	444	1	GUNW_ERMCA
33	98	7.3	922	1	PMPI_CHLPN

34	97.5	7.3	541	1	NU57_YEAST
35	95	7.1	1585	1	YOBO_BACSU
36	90	6.7	526	1	KICJ_BOVIN
37	90	6.7	1161	1	DANA_YEAST
38	90	6.7	1332	1	KKDO_BACSU
39	88	6.6	378	1	SLG1_YEAST
40	88	6.6	1077	1	HLES_DROME
41	87.5	6.6	1061	1	OAR_MYXXA
42	86.5	6.5	72	1	OSMB_SALTY
43	86.5	6.5	1396	1	VITE_BPR5
44	86.5	6.5	3770	1	ACVS_EMENT
45	86	6.4	526	1	VP5_BTIVL

ALIGNMENTS

RESULT 1	CBPA_CLOCL	STANDARD:	PRT:	1848 AA.
ID	CBPA_CLOCL			
AC	P38058:			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	Cellulose binding protein A precursor.			
GN	CBPA.			
OS	Clostridium cellulovorans.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_Taxid=1493;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92228810; PubMed=1555642;			
RA	Shoseyov O., Takagi M., Goldstein M.A., Doi R.H.;			
RT	"Primary sequence analysis of Clostridium cellulovorans cellulose			
RL	binding protein A.";			
Proc. Natl. Acad. Sci. U.S.A.	89:3483-3487(1992).			
CC	-1- FUNCTION: BINDS TO CELLULOSE FIBERS AND COORDINATES CELLULOSE			
CC	ENZYMES.			
CC	-1- PTM: THE N-TERMINUS IS BLOCKED.			
CC	-1- PTM: CONTAINS ABOUT 10% OF CARBOHYDRATES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: M73817; AAA23218.1; -.			
DR	PIR: A44140; A44140.			
DR	HSSP: Q06851; INBC.			
DR	InterPro: IPR001956; CBD_3.			
DR	InterPro: IPR002102; Cohesin.			
DR	Pfam: PF00942; CBD_3; 1.			
DR	Pfam: PF00963; Cohesin; 9.			
DR	ProDom: PD001947; CBD_3; 1.			
KW	Cellulose degradation; Cell wall; Repeat; Signal; Glycoprotein.			
FT	SIGNAL			
FT	CHAIN 29 1848			
FT	DOMAIN 29 189			
SO	SEQUENCE 1848 AA; 189152 MW; 85FA6CE6F771AF1A CRC64;			
Query Match	36.5%; Score 487; DB 1; Length 1848;			
Best local similarity	46.8%; Pred. No. 6.4e-32;			
Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;				
Oy	1 MSVEFYNSKNSAQTNSITPIIKIRNTSDNLNDVKRYRYTSDGTGQGFQFCDHAGAL 60			
Db	33 MSVEFYNSKNSAQTNSITPIIKIRNTSDNLNDVKRYRYTSDGTGQGFQFCDHAGAL 92			
Oy	61 LGNSVDNTSKVTANFVKETASPTSTYDYLDPDSHMRCLGSSLIITSVFLVCAQNF 120			

```

Db 93 LGNSYVONTSKVIANFEKETAAPTSTYDYVE-----124
QY 121 ROEYGAATGAVGVAGVAGLFGKSGRYSMAIGAVLGLGSKIGSQMDQDKITLNSL 180
Db 125 -----FGFASGRATL-----143
QY 181 EKVRAGQVTR--WRNPDTGNSYVEPVRYQRIKQKRRQYQYERFQOKAMTAGROKEIY 238
Db 144 -----GRITKSDMSWYQTNDYSPDASSSTPYVNP-----KVGYIGGAK--VL 185
QY 239 GTACPOPD 246
Db 186 GTA-PCPD 192

```

RESULT 2

```

17KD_RICPR STANDARD; PRT; 159 AA.
ID 17KD_RICPR
AC P16624;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 17 kDa surface antigen precursor.
GN OMP OR RP833.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxId=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE=69359171; PubMed=2768201;
RA Anderson B.E., Tzianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen
gene."
RT J. Bacteriol. 171:5199-5201(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Anderson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
mitochondria."
RT Nature 396:133-140(1998).
RL -1 SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M28482; AAA26378.1; ALT-SEQ.
DR EMBL: AJ235273; CAI15258.1; -.
DR PIR: D33971; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT SEQUENCE 159 AA; 16672 MW; A33D04H65EEB071 CRC64;

```

Query Match 21.6%; Score 288; DB 1; Length 159;
 Best Local Similarity 37.9%; Pred. No. 7.2e-17;
 Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;

QY 103 SSIIISV---FLVGC--AQNFSROEVAATGAVGVAGLFGKSGRYSMAIG-GAVL 156

```

Db 5 SKMITALASMLQACNCGSGMKNKGTGTLGAGGALLSGQFQGGOL-VGVGVGALL 63
QY 157 GGLISKTIQSQMDQDK-----TKLNSLEKVRAGQVTRWRNPDTGNSYVEPVRYQRI 212
Db 64 GAVLGGQIGAGMDEODRRLLETSQRALESAPSSGIMWRPNMGNHGYVPNKTY-----119
QY 213 KOERROQYCREFOQKAMTAGROKEIYGTACPOPDGRMOYIS 253
Db 120 -RNSAGYCREYQTGYVIGKQOKRTYGNACROPDGMOMOVN 159

```

RESULT 3

```

17KD_RICJA STANDARD; PRT; 159 AA.
ID 17KD_RICJA
AC Q52764;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 17 kDa surface antigen precursor.
GN OMP.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxId=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JH;
RX MEDLINE=95229950; PubMed=7714214;
RA Furuya Y., Katayama T., Yoshida Y., Kaiho I.;
RT "Specific amplification of Rickettsia japonica DNA from clinical
specimens by PCR."
RT J. Clin. Microbiol. 33:487-489(1995).
RL -1 SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D16515; BAA03965.1; -.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT SEQUENCE 159 AA; 16554 MW; CDDE7CBBDD6B41 CRC64;

```

Query Match 21.4%; Score 286; DB 1; Length 159;
 Best Local Similarity 37.9%; Pred. No. 1.1e-16;
 Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;

QY 103 SSIIISV---FLVGC--AQNFSROEVAATGAVGVAGLFGKSGRYSMAIG-GAVL 156
 Db 5 SKMITALASMLQACNCGSGMKNKGTGTLGAGGALLSGQFQGGOL-VGVGVGALL 63
 QY 157 GGLISKTIQSQMDQDK-----TKLNSLEKVRAGQVTRWRNPDTGNSYVEPVRYQRI 212
 Db 64 GAVLGGQIGAGMDEODRRLLETSQRALESAPSSGIMWRPNMGNHGYVPNKTY-----119
 QY 213 KOERROQYCREFOQKAMTAGROKEIYGTACPOPDGRMOYIS 253
 Db 120 -RNSAGYCREYQTGYVIGKQOKRTYGNACROPDGMOMOVN 159

RESULT 4
 17KD_RICCN STANDARD; PRT; 159 AA.
 ID 17KD_RICCN
 AC P05372;


```

DT 01-NOV-1988 (Rel. 09, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 17 kDa surface antigen precursor.
GN OMP OR RC1287.
OS Rickettsia conorii, and
OC Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiense; Rickettsia.
OX NCBI_TaxId=781, 783;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=R.conorii, and R.rickettsii;
RX MEDLINE=89359171; PubMed=2768201;
RT Anderson B.E., Tzianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen
RT gene.";
RL J. Bacteriol. 171:5199-5201(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=R.conorii; STRAIN=Wallsh 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Andiffren P., Fournier P.-E., Barbe V.,
RA Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Reault D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RT Science 293:2093-2098(2001).
RL [3]
RP SEQUENCE FROM N.A.
RC SPECIES=R.rickettsii;
RX MEDLINE=87222152; PubMed=3108232;
RA Anderson B.E., Regnery R.L., Carlone G.M., Tzianabos T., McDade J.E.,
RA Fu Z.Y., Bellini W.J.;
RT "Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia
RT rickettsii.";
RL J. Bacteriol. 169:2385-2390(1987).
RN [4]
RP SEQUENCE OF 1-30 FROM N.A.
RC SPECIES=R.rickettsii;
RX MEDLINE=89008059; PubMed=3139629;
RA Anderson B.E., Baunstarck B.R., Bellini W.J.;
RT "Expression of the gene encoding the 17-kilodalton antigen from
RT Rickettsia rickettsii: transcription and posttranslational
RT modification.";
RL J. Bacteriol. 170:4493-4500(1988).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M28479; AAA26379.1; -
DR EMBL; M28480; AAA26376.1; -
DR EMBL; AE008675; AAL03825.1; -
DR EMBL; M16486; AAA26381.1; -
DR EMBL; J03371; -, NOT_ANNOTATED_CDS.
DR PIR; A25972; A25972.
DR PIR; A31836; A31836.
DR PIR; A33971; A33971.
DR PIR; B33971; B33971.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN_1.
KM Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT CONFLICT 146 146 N -> D (IN REF. 3).
FT CONFLICT 153 153 G -> E (IN REF. 3).
SQ SEQUENCE 159 AA: 16561 MW: 20042BB74FC169 CRC64:

```

[illegible]

RESULT 6
 17KD_RICAU STANDARD: PRT: 154 AA.
 ID 17KD_RICAU
 AC P50928:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 17 kDa surface antigen precursor (Fragment).
 GN OMP.
 OS Rickettsia australis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 CC NCBI_TaxID=787;
 RN (1)
 RA Sequence FROM N.A.
 RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 anchor (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: M74042: AAA26394.1;
 DR PROSITE: PS00013: PROKAR_LIPOPROTEIN; 1.
 KW Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
 FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
 FT NON_TER 154 154
 SO SEQUENCE 154 AA; 15967 MW; E3AA833346FAC320 CRC64;
 Query Match 20.1%; Score 268; DB 1; Length 154;
 Best Local Similarity 37.8%; Pred. No. 3e-15;
 Matches 59; Conservative 29; Mismatches 52; Indels 16; Gaps 6;
 OY 103 SSLIIIV---FLVGCAC--NFSRQEVGAATGAVVGAGQLGKSGGRYSMAIG-GAVL 156
 DB 5 SKIMTIALAASMLQACNPGMKNKOGTGTLLGAGGALLSQRKKGQL-VGVGALL 63
 OY 157 GGLISKIGOSMDQDK-----IKLNQSLKVKAGQVTRMNPDTGNSYSEPVRTQRYN 212
 DB 64 GAVLGQIGAGMDEDRRLAELTSGRALETAPSGSNVEMRNDMGNGYVTPNKTYNSN 123
 OY 213 KQRRQOYCREFOQKAMTAGKQKQETIYGTACPODGR 248
 DB 124 -----GQYCREYOTVIVIGKQOKAYGNACROPDQ 154
 RESULT 7
 17KD_RICPA STANDARD: PRT: 154 AA.
 ID 17KD_RICPA
 AC P50930:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 17 kDa surface antigen precursor (Fragment).
 GN OMP.
 OS Rickettsia parkeri.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 CC NCBI_TaxID=35792;
 RN (1)
 RA Sequence FROM N.A.
 RL STRAIN=MACULATUM;
 RC Pretzman C.I., Stochard D.R., Ralph D., Clark J.B., Fuerst P.A.;

RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U11020: AAA2040.1;
 DR PROSITE: PS00013: PROKAR_LIPOPROTEIN; 1.
 KW Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
 FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
 FT NON_TER 154 154
 SO SEQUENCE 154 AA; 15897 MW; 5D06F45F9BDE5EC CRC64;
 Query Match 19.4%; Score 259; DB 1; Length 154;
 Best Local Similarity 36.5%; Pred. No. 1.7e-14;
 Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;
 OY 103 SSLIIIV---FLVGC--AQNFSRQEVGAATGAVVGAGQLGKSGGRYSMAIG-GAVL 156
 DB 5 SKIMTIALAASMLQACNPGMKNKOGTGTLLGAGGALLSQRKKGQL-VGVGALL 63
 OY 157 GGLISKIGOSMDQDK-----IKLNQSLKVKAGQVTRMNPDTGNSYSEPVRTQRYN 212
 DB 64 GAVLGQIGAGMDEDRRLAELTSGRALETAPSGSNVEMRNDMGNGYVTPNKTYN 119
 OY 213 KQRRQOYCREFOQKAMTAGKQKQETIYGTACPODGR 248
 DB 120 -RSTGQYCREYOTVIVIGKQOKAYGNACROPDQ 154
 RESULT 8
 17KD_RICRH STANDARD: PRT: 154 AA.
 ID 17KD_RICRH
 AC P50931:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 17 kDa surface antigen precursor (Fragment).
 GN OMP.
 OS Rickettsia rhipicephali.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 CC NCBI_TaxID=31992;
 RN (1)
 RA Sequence FROM N.A.
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U11020: AAB0706.1;
 DR PROSITE: PS00013: PROKAR_LIPOPROTEIN; 1.
 KW Outer membrane; Lipoprotein; Antigen; Signal.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
 FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
 FT NON_TER 154 154

SEQUENCE 154 AA: 15895 MW: 0CF85AD5D96DFE6B CRC64;

Query Match 19.4%; Score 259; DB 1; Length 154;

Best Local Similarity 36.5%; Pred. No. 1.7e-14; Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

103 SLLIISV---FLVGC--AQNFSRQEVGATGAVVGGVAGQLPFKSGSRVSMATG-GAVL 156

5 SKIMITIALASMLQACNGPGGMKQGTGTLGAGGALLGSQFQKGGQL-VGVGVGALL 63

157 GGLIGSKIGSMDQDK---IKLNSLEKVKAGQVTRMRNPDGNSVSVEPVRYORYN 212

64 GAVLGQIGAGMDQDRRLAELTYSQALETAPSGSNVEMRNPDNGNYITPNKTY----- 119

213 KOERROOYCREFOOKAMIAQOKETITGTACPOPDR 248

120 -RNSTGQYCREYTOTVYIGKQKAYGNACLOPDQO 154

RESULT 9

17KD_RICMO STANDARD: PRT; 154 AA.

AC P50929;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE 17 Kda surface antigen precursor (Fragment).

GN OMP.

OS Rickettsia montana.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=33991;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OHIO 83-441;

RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;

RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor (Probable).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U11017; AAB07705.1; -

DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.

KW Outer membrane; Lipoprotein; Antigen; Signal.

FT SIGNAL 1 19 BY SIMILARITY.

FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.

FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).

FT NON_TER 154 154

SEQUENCE 154 AA: 15881 MW: A09C53B8769E31DA CRC64;

Query Match 19.2%; Score 256; DB 1; Length 154;

Best Local Similarity 36.5%; Pred. No. 2.9e-14; Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

103 SLLIISV---FLVGC--AQNFSRQEVGATGAVVGGVAGQLPFKSGSRVSMATG-GAVL 156

5 SKIMITIALASMLQACNGPGGMKQGTGTLGAGGALLGSQFQKGGQL-VGVGVGALL 63

157 GGLIGSKIGSMDQDK---IKLNSLEKVKAGQVTRMRNPDGNSVSVEPVRYORYN 212

64 GAVLGQIGAGMDQDRRLAELTYSQALETAPSGSNVEMRNPDNGNYITPNKTY----- 119

213 KOERROOYCREFOOKAMIAQOKETITGTACPOPDR 248

120 -RNSTGQYCREYTOTVYIGKQKAYGNACLOPDQO 154

RESULT 10

17KD_RICAM STANDARD: PRT; 154 AA.

AC P50927;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE 17 Kda surface antigen precursor (Fragment).

GN OMP.

OS Rickettsia amblyomni.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=33989;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MO 85-1084;

RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;

RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor (Probable).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U11013; AAB07704.1; -

DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.

KW Outer membrane; Lipoprotein; Antigen; Signal.

FT SIGNAL 1 19 BY SIMILARITY.

FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.

FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).

FT NON_TER 154 154

SEQUENCE 154 AA: 15879 MW: E4FB64C9D943581 CRC64;

Query Match 19.1%; Score 255; DB 1; Length 154;

Best Local Similarity 36.5%; Pred. No. 3.5e-14; Matches 57; Conservative 29; Mismatches 54; Indels 16; Gaps 6;

103 SLLIISV---FLVGC--AQNFSRQEVGATGAVVGGVAGQLPFKSGSRVSMATG-GAVL 156

5 SKIMITIALASTLQACNGPGGMKQGTGTLGAGGALLGSQFQKGGQL-VGVGVGALL 63

157 GGLIGSKIGSMDQDK---IKLNSLEKVKAGQVTRMRNPDGNSVSVEPVRYORYN 212

64 GAVLGQIGAGMDQDRRLAELTYSQALETAPSGSNVEMRNPDNGNYITPNKTY----- 119

213 KOERROOYCREFOOKAMIAQOKETITGTACPOPDR 248

120 -RNSTGQYCREYTOTVYIGKQKAYGNACLOPDQO 154

RESULT 11

CIPB_CLOTM STANDARD: PRT; 772 AA.

AC 001866;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein SL/SL)

GN CIPB.

OS Clostridium thermocellum.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1515;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=YS.

```

RX MEDLINE-93146373; PubMed-1490597;
RA Gilbert D.M., Morag E., Lamed R., Bayer E.A., Hazlewood G.P.,
RT "Identification of the cellulose-binding domain of the cellulosome
RL subunit S1 from Clostridium thermocellum vs. ";
CC FEMS Microbiol. Lett. 78:181-186(1992).
CC -1- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
CC CELLULOLYTIC ENZYMES PROBABLY THROUGH THE BINDING OF THE NINE
CC REPEATED DOMAINS WITH THE DUPLICATED SEGMENT OF 24 AMINO ACIDS
CC PRESENT IN CATALYTIC SUBUNITS OF THE CELLULOSE.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE.
CC -1- SIMILARITY: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X68233; CAA48312.1; -.
DR HSSP: 006851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR002102; Cohesin.
DR InterPro: IPR002105; Dockerin_1.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00963; Cohesin; 3.
DR Pfam: PF00404; Dockerin_1; 2.
DR PRODOM: PD001947; CBD_3; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS00448; CLOS_CELLULOSE_RPT; 2.
DR PROSITE: PS00448; CLOS_CELLULOSE_RPT; 2.
KW Cellulose degradation; Cell wall; Glycoprotein; Repeat.
FT NON_TER 1
FT DOMAIN 1 607 3 X 147 AA APPROXIMATE REPEATS.
FT REPEAT 1 607 1 (PARTIAL).
FT REPEAT 2 240 2.
FT REPEAT 3 462 607 3.
FT DOMAIN 81 93 PRO/THR-RICH.
FT DOMAIN 241 272 PRO/THR-RICH.
FT DOMAIN 273 439 CELLULOSE-BINDING.
FT DOMAIN 440 461 PRO/THR-RICH.
FT DOMAIN 710 766 2 X 24 AA APPROXIMATE REPEATS.
FT REPEAT 710 733 2-1.
FT REPEAT 743 766 2-2.
SO SEQUENCE 772 AA; 82491 MW; BBF06DE5E094FE10 CRC64;

Query Match 18.4%; Score 245; DB 1; Length 772;
Best Local Similarity 51.5%; Pred No. 1.5e-12;
Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;

QY 1 MSVEEYNSNKSQNTSIPRIKINTSDSDINLDVKVRYIYSDGQGTFCMCDHAGAL 60
DB 281 LKVEEYNSNPTNSINPQKRYVTGSSAIDSLKTLRYIYVDDGKQDTFMDHA-AT 339
QY 61 LCN--SYVDNISKYVANTFVKETASPTSYDYIDPSHMRGLQ 101
DB 340 IGSNGSYNGITNSVKGTFVKMSSS-TNNADYILEISFTGILE 381

RESULT 12
CIPA_CLOTM STANDARD: PR1; 1853 AA.
AC 006851;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cellulosomal scaffolding protein A precursor (cellulosomal
DE glycoprotein S1/S1) (Cellulose integrating protein A) (Cohesin).
CIPA.
OS Clostridium thermocellum.

```

```

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-40 AND 42-43.
RC STRAIN-ATCC 27405;
RX MEDLINE-93302508; PubMed-8316083;
RA Gerngross U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S.,
RA Demain A.L.;
RT "Sequencing of a Clostridium thermocellum gene (cipa) encoding the
RT cellulosomal SL-protein reveals an unusual degree of internal
RT homology.";
RL Mol. Microbiol. 8:325-334(1993).
RN [2]
RP SEQUENCE OF 1820-1853 FROM N.A.
RX MEDLINE-93209931; PubMed-8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein Cipa and a protein possibly involved
RT in attachment of the cellulosome to the cell surface.";
RL J. Bacteriol. 175:1891-1899(1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.
RX MEDLINE-97238934; PubMed-9083107;
RA Shimon L.J.W., Bayer E.A., Morag E., Yaron S., Shoham Y.,
RA Frolov F.;
RT "A cohesin domain from Clostridium thermocellum: the crystal
RT structure provides new insights into cellulosome assembly.";
RL Structure 5:381-390(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
RX MEDLINE-97076134; PubMed-8918451;
RA Torno J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,
RA Steltz T.A.;
RT "Crystal structure of a bacterial family-III cellulose-binding
RT domain: a general mechanism for attachment to cellulose.";
RL EMBO J. 15:5739-5751(1996).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
RX MEDLINE-98022914; PubMed-9402065;
RA Tavares G.A., Beguin P., Alzari P.M.;
RT "The crystal structure of a type I cohesin domain at 1.7-A
RT resolution.";
RL J. Mol. Biol. 273:701-713(1997).
CC -1- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
CC CELLULOLYTIC ENZYMES.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE.
CC -1- DOMAIN: CONTAINS 9 COPIES OF A DOMAIN (COHESIN) OF ABOUT 150
CC RESIDUES THAT BINDS TO THE DOCKERIN DOMAIN BORN BY THE CATALYTIC
CC COMPONENTS OF THE CELLULOSE.
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L08665; -. NOT_ANNOTATED_CDS.
DR EMBL: X67506; CAA47840.1; -.
DR PIR: S36859; S36859.
DR PDB: 1ANU; 23-JUL-97.
DR PDB: 1AOK; 08-JUL-98.
DR PDB: INBC; 26-SEP-97.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR002102; Cohesin.
DR InterPro: IPR002105; Dockerin_1.
DR InterPro: IPR002048; EF-hand.

```

DR Pfam: PF00942; CBD_3; 1.
 DR Pfam: PF00963; Cohesin; 9.
 DR Pfam: PF00404; Dockerin; 2.
 DR ProDom: PD001947; CBD_3; 1.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN; 1.
 DR PROSITE: PS00448; CLOS_CELLOSOME_RPT; 2.
 KW Cellulose degradation; Cell wall; Glycoprotein; Repeat; Signal;
 3D-structure.
 FT SIGNAL 1 28
 FT CHAIN 29 1853
 FT DOMAIN 364 522
 FT DOMAIN 54 1694
 FT REPEAT 29 182
 FT REPEAT 183 322
 FT REPEAT 560 704
 FT REPEAT 724 866
 FT REPEAT 889 1031
 FT REPEAT 1054 1196
 FT REPEAT 1219 1361
 FT REPEAT 1384 1526
 FT REPEAT 1548 1690
 FT DOMAIN 1791 1846
 FT REPEAT 1791 1814
 FT REPEAT 1824 1847
 FT REPEAT 1615 1615
 FT CONFLICT 1615 1615
 SQ SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABFC5372 CRC64;

Query Match 18.48; Score 245; DB 1; Length 1853;
 Best Local Similarity 51.58; Pred. No. 4.1e-12;
 Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;

QY 1 MSVPEYNSKSAQNTSITPIKIKINTSDSLNDVKKVRYRYSDCGTGCTGCFWCDHAGAL 60
 Db 369 LKVEFYNSNPDTNTSINPQKVTNTGSSALDSLTLRYTYVDGQKDTFWCDHA-AI 427
 61 LGN--SYVDNTSKVTANFKETASPTSTYDTYLDPSHMRGCLQ 101
 Db 428 IGSNGSYNGITSNKGTFVKMSS--TNNADYLEISFTGGLLE 469

RESULT 13
 GUN1_CLOTM STANDARD; PRT; 879 AA.
 AC 002934;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Endoglucanase I precursor (EC 3.2.1.4) (Egri) (Endo-1,4-beta-glucanase)
 DE (Cellulase I).
 CN CELL.
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_Taxid=1515;
 RN [1].
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.
 RC STRAIN=NCIB 10662; PubMed=8436949;
 RA MEDLINE=93171873; PubMed=8436949;
 RA Haelewood G.P., Davidson K., Laurie J.I., Huskisson N.S.,
 RA Gilbert H.J.;
 RT "Gene sequence and properties of Cell, a family E endoglucanase from
 Clostridium thermocellum";
 RL J. Gen. Microbiol. 139:307-316(1993).
 CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
 GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
 GLUCANS. PRINCIPALLY ACTIVE AGAINST BARLEY BETA-GLUCAN.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 linkages in cellulose.
 CC -1- PATHWAY: CELLULOSE DEGRADATION.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL
 HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: I04735; AAA20892.1; -.
 DR PIR: A47704; A47704.
 DR HSSP: P26221; ITR4.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR001701; Glyco_hydro_9.
 DR Pfam: PF00942; CBD_3; 2.
 DR Pfam: PF00759; Glyco_hydro_9; 1.
 DR ProDom: PD001947; CBD_3; 1.
 DR PROSITE: PS00592; GLYCOSYL_HYDROL_P9_1; 1.
 DR PROSITE: PS00698; GLYCOSYL_HYDROL_P9_2; 1.
 KW Cellulose degradation; Hydrolyase; Glycosidase; Signal.
 FT SIGNAL 1 55
 FT CHAIN 56 879
 FT DOMAIN 56 518
 FT DOMAIN 729 879
 FT ACT_SITE 448 448
 FT ACT_SITE 486 486
 FT ACT_SITE 495 495
 FT ACT_SITE 495 495
 SQ SEQUENCE 879 AA; 97796 MW; 35A60069A514A927 CRC64;

Query Match 14.18; Score 188.5; DB 1; Length 879;
 Best Local Similarity 28.48; Pred. No. 7.3e-08;
 Matches 59; Conservative 20; Mismatches 48; Indels 81; Gaps 9;

QY 3 VEFYNSKSAQNTSITPIKIKINTSDSLNDVKKVRYRYSDCGTGCTGCFWCDHAGALLG 62
 Db 742 LQYANGNAGATSNISNPFKIKINGTKAINDVKIRIYTKEGASQNFCDHSSA--G 799
 63 NSYVDNTSKVTANFKETASPTSTYDTYLDPSHMRGCLQSSLIITISVFLGCAQNFESRQ 122
 Db 800 N-----SNVTGNF-NLSSPKEGADT-----CL----- 821

QY 123 EVGAATGAVVGAGQLFGKSGRYSMAGAVLGLGSKIGOSMDQDKIKINSLER 182
 Db 822 EVG-----FGSGAG--TLDPGGSV-----EVQIRFSKED 848
 QY 183 VKAGQVTRMRPDPGNSVSEVPRTYR 210
 Db 849 -----WSNYNOSNDYSFKOACLQR 868

RESULT 14
 GUN2_CLOSR STANDARD; PRT; 986 AA.
 AC P23659;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Endoglucanase Z precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Thermocactive cellulase) (Avicelase I).
 CN CELL.
 OS Clostridium stercorarium.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_Taxid=1510;
 RN [1].
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-36 AND 475-486.
 RC STRAIN=NCIB 11745;
 RA MEDLINE=91066838; PubMed=2250652;
 RA Jauris S., Ruecknagel K.P., Schwarz W.H., Kratzsch P.,
 RA Bronnenmeier K., Staudenbauer W.L.;
 RT "Sequence analysis of the Clostridium stercorarium celZ gene encoding
 a thermocactive cellulase (Avicelase I): Identification of catalytic
 RT and cellulose-binding domains";
 RL Mol. Gen. Genet. 223:258-267(1990).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic

linkages in cellulose.
 CC -1- SIMILARITY: CONTAINS MULTIPLE CELLULOSE-BINDING SITES.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X55299; CAA39010.1; ALT_SEQ.
 DR PIR: S12021; S12021.
 DR HSSP: P26221; 1TF4.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR001701; Glyco_hydro_9.
 DR Pfam: PF00942; CBD_3; 2.
 DR Pfam: PF00759; Glyco_hydro_9; 1.
 DR ProDom: PD001947; CBD_3; 1.
 DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
 DR PROSITE: PS00598; GLYCOSYL_HYDROL_F9_2; 1.
 DR Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
 KW SIGNAL 1 25
 FT CHAIN 26 986 ENDOLUCANASE 2.
 FT REPEAT 498 641 DOMAIN C'.
 FT REPEAT 651 738 DOMAIN B.
 FT REPEAT 744 831 DOMAIN B.
 FT REPEAT 854 986 DOMAIN C.
 FT DOMAIN 835 986 CELLULOSE-BINDING (BY SIMILARITY).
 FT ACT_SITE 400 400 BY SIMILARITY.
 FT ACT_SITE 438 438 BY SIMILARITY.
 FT ACT_SITE 447 447 BY SIMILARITY.
 SQ SEQUENCE 986 AA: 109512 MW: 180280982923690 CRC64;
 Query Match 14.1%; Score 188; DB 1; Length 986;
 Best Local Similarity 27.6%; Pred. No. 9.2e-08;
 Matches 47; Conservative 29; Mismatches 58; Indels 36; Gaps 6;
 QY 1 MSVFYNSKSAQNTSTPIIKITNTSDSLNDVRYRYTSDGTGQTFWCHAGAL 60
 Db 840 IOIDFNGNSTDKNITMPKRLNCTPIRLSDVKIRYYTIDEDKDNFCWSSV- 898
 QY 61 LGNSYDNTSKVTANFYKETASPTSTYDTYLDSSHARGC--LQSSLLIISVPLVCAON 118
 Db 899 -----GSSNITGTFFVK-MAEPKEGADYLETGFTDGAGTLOPNOSIEVQ-----NR 943
 QY 119 FSR-----QEVGATGAVGVGAGVAGQLFGKSGRVSMAIGAVLGL 159
 Db 944 FSKADMTDIOTNDYSTSTNTSYG-----SNDRIITYISGVAVSGI 984
 RESULT 15
 GUX2_CLOSR STANDARD: PRT; 914 AA.
 ID GUX2_CLOSR
 AC P50900;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 01-NOV-1997 (rel. 35, Last annotation update)
 DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellolbiohydrolase II)
 DE (1,4-beta-cellobiohydrolase II) (Avicelase II).
 GN CELV.
 OS Clostridium stercorarium.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1510;
 RN [1]
 RP SOURCE FROM N.A.
 RC STRAIN-NCIB 11754;
 RA Staudenbauer W.L., Kundt K., Riedel K., Schwarz W.H.,
 RL Submitted (FEB-1996) to the EMBL/genbank/DBJ databases.

RN [2]
 RP CHARACTERIZATION.
 RC STRAIN-NCIB 11754;
 RX MEDLINE-91364686; PubMed-1909625; Staudenbauer W.L.;
 RA Bromenneier K., Ruecknagel K.P., Staudenbauer W.L.;
 RT "Purification and properties of a novel type of
 RT exo-1,4-beta-glucanase (avicelase II) from the cellulolytic
 RT thermophile Clostridium stercorarium";
 RL Eur. J. Biochem. 200; 379-385(1991).
 CC -1- CATALYTIC ACTIVITY: Hydrolase of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellotetraose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY L (FAMILY 48 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z69359; CAA93280.1; -
 DR HSSP: O06851; INBC.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR005556; Glyco_hydro_48.
 DR Pfam: PF00942; CBD_3; 1.
 DR Pfam: PF00844; GLHYDRLASE48.
 DR PRINTS: PR00844; GLHYDRLASE48.
 DR ProDom: PD001947; CBD_3; 1.
 DR ProDom: PD011903; Glyco_hydro_48; 1.
 DR Cellulose degradation; Hydrolase; Glycosidase; Signal.
 KW SIGNAL 1 33
 FT CHAIN 34 914 EXOGLUCANASE II.
 FT SEQUENCE 914 AA: 103020 MW: D0DB6017D6DFF82C CRC64;
 Query Match 12.2%; Score 162.5; DB 1; Length 914;
 Best Local Similarity 37.8%; Pred. No. 1e-05;
 Matches 34; Conservative 17; Mismatches 30; Indels 9; Gaps 2;
 QY 3 VEFYNSKSAQNTSTPIIKITNTSDSLNDVRYRYTSDGTGQTFWCHAGALLG 62
 Db 769 IQSNFATQETISNSIMPRFRIYNSGNTSIPLSVKKLRYTYVDGKPKQNFQCDNASI-- 825
 QY 63 NSYDNTSKVTANFYKETASPTSTYDTYLD 92
 Db 826 -----GSSNVTGTFFVKMDGATIGA-DYILE 849
 Search completed: October 27, 2002, 11:10:53
 Job time : 11.7271 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 11:05:56 ; Search time 16.3592 Seconds
(without alignments)
1503.668 Million cell updates/sec

Title: US-09-677-374-6

Perfect score: 1335
Sequence: 1 MSVEFYNSNKSAGTNSITPI.....ITGTACPDQPDGRWQVISTEK 256

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:1*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	487	36.5	1848	2 A44140	cellulose-binding
2	288	21.6	159	2 D33971	ricketsial common
3	285	21.3	159	2 B33971	ricketsial common
4	285	21.3	159	2 A33971	ricketsial common
5	285	21.3	159	2 G97860	17K surface antige
6	276.5	20.7	159	2 C33971	ricketsial common
7	276	20.7	159	2 A25972	17K antigen precu
8	249	18.7	586	2 PC6006	scaffolding protei
9	245	18.4	1854	2 S36859	cipa protein - Clo
10	233.5	17.5	1162	2 T30433	scaffolding protei
11	223.5	16.7	1483	2 G47012	probably cellulose
12	188.5	14.1	879	2 A47704	endoglucanase I (E
13	188	14.1	986	2 S12021	thermoactive cellu
14	151.5	11.3	1779	2 T31085	xylanase - Caldice
15	147	11.0	915	2 A43802	cellulase (EC 3.2.
16	143.5	10.7	1039	2 S02711	cellulase (EC 3.2.
17	143.5	10.7	1331	2 A48954	mannan endo-1,4-be
18	142	10.6	1711	2 T31337	1,4-beta-glucanase
19	138.5	10.4	1742	2 T17120	cellulase (EC 3.2.
20	127.5	9.6	182	2 D83169	conserved hypotet
21	122	9.1	257	2 F87413	hypothetical prote
22	121	9.1	508	2 G69593	cellulase (EC 3.2.
23	120.5	9.0	131	2 AI3418	17K surface antige
24	115	8.6	499	2 JN0111	cellulase (EC 3.2.
25	113.5	8.5	499	2 A27198	cellulase (EC 3.2.
26	112.5	8.4	155	2 S23787	outer membrane lip
27	111.5	8.4	154	2 B83514	conserved hypotet
28	111	8.3	145	2 A41897	cellulase homology
29	109.5	8.2	508	2 A26874	cellulase (EC 3.2.

30	109	8.2	155	2 A10693	outer membrane lip
31	109	8.2	179	2 AE0644	probable secreted
32	109	8.2	232	2 G87629	hypothetical prote
33	109	8.2	257	2 B82837	conserved hypotet
34	107.5	8.1	142	2 AD2696	lipa protein (limpo
35	107.5	8.1	155	2 AF0289	probable lipoprote
36	107	8.0	125	2 D97438	lipa protein (limpo
37	107	8.0	486	2 I40548	bifunctional cellu
38	106.5	8.0	155	2 AG0443	outer membrane lip
39	104.5	7.8	155	2 C64921	outer membrane lip
40	104.5	7.8	155	2 F90922	probable outer mem
41	104.5	7.8	139	2 B85771	probable outer mem
42	104	7.8	139	2 S58234	lipa protein - Rhl
43	103	7.7	232	2 AD3350	outer membrane pro
44	103	7.7	504	2 S54744	cellulase (EC 3.2.
45	101.5	7.6	155	2 I64130	PAL cross-reacting

ALIGNMENTS

RESULT 1
A44140
cellulose-binding protein A precursor - Clostridium cellulovorans
C:Species: Clostridium cellulovorans
C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 15-Oct-1999
C:Accession: A44140
R:Shoseyov, O.; Takagi, M.; Goldstein, M.A.; Doi, R.H.
Proc. Natl. Acad. Sci. U.S.A. 89, 3483-3487, 1992
A:Title: Primary sequence analysis of Clostridium cellulovorans cellulose binding pro
A:Reference number: A44140; MUID:92228810
A:Accession: A44140
A>Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-1848 <SHO>
A:Cross-references: GB:M73817; NID:g144748; PIDN:AAA23218.1; PID:g144749
C:Genetics:
A:Gene: cbpa

Query Match 36.5%; Score 487; DB 2; Length 1848;
Best Local Similarity 46.8%; Pred. No. 2.8e-32;
Matches 116; Conservativity 12; Mismatches 30; Indels 90; Gaps 7;

QY	1	MSVEFYNSNKSAGTNSITPIIKITNTSDSLNLDVAVRYRYTSDGQGTFCMDHAGAL	60
DB	33	MSVEFYNSNKSAGTNSITPIIKITNTSDSLNLDVAVRYRYTSDGQGTFCMDHAGAL	92
QY	61	LGNSYVDNTSKYANFYKERTASPTSTYDTYLDPSHMRGCLQGSLLIISVFLVGAQNF	120
DB	93	LGNSYVDNTSKYANFYKERTASPTSTYDTYLDPSHMRGCLQGSLLIISVFLVGAQNF	124
QY	121	ROEYGAATGAVGVAGQGLFGKSGRVSMAIGAVLGLIGSKIGQMDQDKIKLNQSL	180
DB	125	-----FGFASGRATL-----KKGQRTIQ-----	143
QY	181	EKKYKAGVTR--WRNPDTGSGSYVEPRYTORYNKQRRQOYCREPQKAMINGQKEIY	238
DB	144	-----GRITKSDMSNTYTQTDYSFDASSRPVNP-----KVTGIGGAK--VL	185
QY	239	GTACPOPD 246	
DB	186	GTA-PGPD 192	

RESULT 2
D33971
ricketsial common antigen precursor - Rickettsia prowazekii
N:Alternate names: 17kD surface antigen, outer membrane protein (omp); RP833
C:Species: Rickettsia prowazekii
C:Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 03-Nov-2000
C:Accession: D33971; B71645
R:Anderson, B.E.; Tzianabos, T.
J. Bacteriol. 171, 5199-5201, 1969

RESULT 8
PC6006
scaffolding protein CtpC precursor - Clostridium cellulolyticum (fragment)
C:Species: Clostridium cellulolyticum
C.Date: 17-Aug-1990 #sequence-revision 16-Aug-1996 #text-change 13-Nov-1998
C.Accession: PC6006
R:Pages, S.; Belaich, A.; Tardif, C.; Reverbel-Leroy, C.; Gaudin, C.; Belaich, J.P.
J. Bacteriol. 178, 2279-2286, 1996

RESULT 9
 536859
 c1pA protein - Clostridium thermocellum
 N:Alternate names: probable cellulosome protein large chain SL
 C:Species: Clostridium thermocellum
 C:Date: 13-Jan-1995 #sequence=revision 13-Jan-1995 #text_change 01-Dec-2000
 C:Accession: S36859; S33527; S25767; S28659; T18261
 R:Gerngross, U.T.; Demain, A.L.
 submitted to the EMBL Data Library, January 1993
 A:Reference number: S36859
 A:Accession: S36859
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1854 <GER>
 A:Cross-references: EMBL:L08665
 R:Gerngross, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.
 Mol. Microbiol. 8, 325-334, 1993
 A>Title: Sequencing of a Clostridium thermocellum gene (c1pA) encoding the cellulosome
 A:Reference number: S33527; MUID:93302508
 A:Accession: S33527
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1615,1617-1854 <GE2>
 A:Cross-references: EMBL:L08665
 R:Fujino, T.; Beguin, P.; Aubert, J.P.
 FEBS Microbiol. Lett. 94, 165-170, 1992
 A>Title: Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides tha
 A:Reference number: S25767
 A:Accession: S25767
 A:Molecule type: DNA
 A:Residues: 1216-1232, 'V', 1234-1241, 'T', 1243-1319, 'R', 1321-1615, 1617-1854 <FUS>
 A:Cross-references: EMBL:X67406
 R:Beugin, P.
 submitted to the EMBL Data Library, August 1992
 A:Reference number: S28659
 A:Accession: S28659

Db 849 -----MSVNFNSNDYSEKQACLROR 868

RESULT 13

S12021

thermoactive cellulase - Clostridium stercorarium
N:Alternate names: avicelase I

C:Species: Clostridium stercorarium

C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997

C:Accession: S12021

R:Authors, S.: Ruecknagel, K.P.; Schwarz, W.H.; Kratzsch, P.; Bronnenmeier, K.; Staudenb

Mol. Gen. Genet. 223, 258-267, 1990

A:Title: Sequence analysis of the Clostridium stercorarium celz gene encoding a thermoac

A:Reference number: S12021; MUID:9106838

A:Accession: S12021

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-986 <JAU>

Query Match 14.1%; Score 188; DB 2; Length 986;
Best Local Similarity 27.6%; Pred. No. 1.4e-07;

Matches 47; Conservative 29; Mismatches 58; Indels 36; Gaps 6;

QY 1 MSVEFYNSKNSAQTNSITPIIKITNTSDSLNDVKVRYVYSDGTQGGTFWCDHAGL 60

DB 840 IQIQMFNGNTSDKTNGLMPRYRLNTGTPIRLSDVKIRYRTIDGKDNQFMCWMSV- 898

QY 61 LGNSVYDNTSKVTANFVKETASPTSTYDYLDPHMRGC--LQSSLLIISVFLVCAON 118

DB 899 -----GSNNITGTFFVK-MAEPKEGADYYLETGFTDAGYLPNQSLIEVO-----NR 943

QY 119 FSR-----DEVGATGAVGVAGOLF-GKSGRVSMAIGAVGGL 159

DB 944 FSKRMDWDTYDTNDYSESTNTSYG-----SNDKITVYISVLVSGI 984

RESULT 14

T31085

xylinase - Caldicellulosiruptor sp.

C:Species: Caldicellulosiruptor sp.

C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

C:Accession: T31085

R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.

submitted to the EMBL data library, December 1997

A:Description: Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. Rt69B.1.

A:Reference number: Z20972

A:Accession: T31085

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1779 <MOR>

A:Cross-references: EMBL:AF036924; NID:g2760905; PID:g2760909; PIDN:AB95326.1

C:Genetics:

A>Note: xync

Query Match 11.3%; Score 151.5; DB 2; Length 1779;

Best Local Similarity 33.6%; Pred. No. 0.00032;

Matches 51; Conservative 15; Mismatches 35; Indels 51; Gaps 6;

QY 1 MSVEFYNSKNSAQTNSITPIIKITNTSDSLNDVKVRYVYSDGTQGGTFWCDHA--G 58

DB 1114 LKVLKNNETTSASGTIRPMFKIVNGSSSVLDSRVKIRIYWTVDGKPGQAVCDMAQIG 1173

QY 59 ALLGNSYVDNTSKVTANFVKETASPTSTYDYLDPHMRGCLQSSLLIISVFLVCAON 118

DB 1174 A-----SNVTFNFVKLT-SGVSGADYLL----- 1195

QY 119 FSRQEVGAATGAVGVAGOLF-GKSGGRVSM 149

DB 1196 -----EVGFSSG-----AGQLQPGKDTGDIQV 1217

RESULT 15

A43802

cellulase (EC 3.2.1.4) / cellulose 1,4-beta-D-glucosidase (EC 3.2.1.91) - Caldocellu

N:Alternate names: endo-1,4-beta-D-glucanase

C:Species: Caldocellum saccharolyticum 30-Sep-1993 #text_change 10-Jul-1998

C>Date: 30-Jan-1993 #sequence_revision 30-Sep-1993

C:Accession: A43802

R:Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P

Appl. Environ. Microbiol. 56, 3117-3124, 1990

A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophil

A:Reference number: A43802; MUID:91136262

A:Accession: A43802

A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual

A:Molecule type: DNA

A:Residues: 1-915 <SAU>

A:Cross-references: EMBL:X13602

C:Function: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F;20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SNV>

Query Match 11.0%; Score 147; DB 2; Length 915;
Best Local Similarity 23.2%; Pred. No. 0.00034;

Matches 60; Conservative 35; Mismatches 64; Indels 100; Gaps 11;

QY 1 MSVEFYNSKNSAQTNSITPIIKITNTSDSLNDVKVRYVYSDGTQGGTFWCDHA--G 58

DB 351 IKVLYANKETNSTNTIRPMFKIVNCGSSSIDLSRVYIRIYWTVDGERAOSAVSDMAQIG 410

QY 59 ALLGNSYVDNTSKVTANFVKETASPTSTYDYLDPHMRGCLQSSLLIISVFLVCAON 118

DB 411 A-----SNVTFNFVK-LSSVSGADYLL----- 432

QY 119 FSRQEVGAATGAVGVAGOLF-GKSGGRVSMALGAVLGCLGSKIGQSMDDQDKIKLN 177

DB 433 -----EIGFKSG-----AGQLQPGKDTGEI-----QIRFN 457

QY 178 OSLEKVKAGVYTRMRNDPTGNSYSEPEYRTQRYNKOERRQOYCRFQOKAMINGOKEI 237

DB 458 KS-----DWSNTNQGNDMSW--LQSMSTYGENEKTATYI---DGVLVNGOE--- 498

QY 238 YGTACPOPDGRWQVISTEK 256

DB 499 -----PSTNDMLVYSGNK 512

Search completed: October 27, 2002, 11:11:39
Job time : 19.3592 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 11:08:02 : Search time 13 2642 Seconds
(without alignments)
471.414 Million cell updates/sec

Title: US-09-677-374-6

Perfect score: 1335
Sequence: 1 MSVEFYNSNKSQNTSITPTI.....ITGTCAPQPDGRNQVISTEK 256

Scoring table:

BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/6C.COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/6D.COMB.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	480.5	36.0	382	4 US-09-277-716-22	Sequence 22, Appl
2	477	35.7	162	1 US-08-048-164A-2	Sequence 2, Appl
3	477	35.7	162	1 US-08-460-462-2	Sequence 2, Appl
4	477	35.7	162	1 US-08-460-457-2	Sequence 2, Appl
5	477	35.7	162	1 US-08-460-458-2	Sequence 2, Appl
6	477	35.7	162	1 US-08-460-455-2	Sequence 2, Appl
7	477	35.7	162	3 US-08-330-394A-2	Sequence 2, Appl
8	477	35.7	163	3 US-09-006-636-7	Sequence 7, Appl
9	477	35.7	163	4 US-09-006-632-7	Sequence 29, Appl
10	472	35.4	154	2 US-08-330-394A-29	Sequence 22, Appl
11	472	35.4	156	2 US-08-330-394A-22	Sequence 22, Appl
12	472	35.4	167	5 PCT-US95-13813-9	Sequence 9, Appl
13	245	18.4	493	4 US-09-198-956-10	Sequence 12, Appl
14	245	18.4	493	4 US-09-198-955A-12	Sequence 12, Appl
15	149.5	11.2	1426	4 US-09-136-574A-43	Sequence 43, Appl
16	148.5	11.1	616	4 US-09-136-574A-47	Sequence 47, Appl
17	148.5	11.1	1751	4 US-09-136-574A-44	Sequence 44, Appl
18	122.5	9.2	551	2 US-09-033-537A-1	Sequence 1, Appl
19	100.5	7.5	700	2 US-07-862-588B-2	Sequence 12, Appl
20	85	6.4	1021	1 US-07-910-760-12	Sequence 12, Appl
21	85	6.4	1021	1 US-08-440-519-12	Sequence 12, Appl
22	85	6.4	1021	1 US-08-440-549-12	Sequence 12, Appl
23	83	6.2	382	4 US-08-444-818-68	Sequence 68, Appl
24	83	6.2	460	4 US-08-444-818-20	Sequence 58, Appl
25	83	6.2	512	4 US-08-867-611-58	Sequence 47, Appl
26	83	6.2	592	4 US-08-867-611-47	Sequence 47, Appl
27	83	6.2	594	4 US-08-867-611-48	Sequence 48, Appl

28	83	6.2	597	4 US-08-867-611-16	Sequence 16, Appl
29	83	6.2	597	5 PCT-US92-06965A-21	Sequence 21, Appl
30	83	6.2	599	4 US-08-867-611-18	Sequence 18, Appl
31	83	6.2	599	5 PCT-US92-06965A-23	Sequence 23, Appl
32	83	6.2	613	4 US-08-867-611-49	Sequence 49, Appl
33	83	6.2	739	4 US-08-444-818-148	Sequence 148, Appl
34	83	6.2	781	4 US-08-867-611-4	Sequence 4, Appl
35	83	6.2	781	5 PCT-US92-06965A-9	Sequence 9, Appl
36	83	6.2	859	1 US-08-444-818-30	Sequence 30, Appl
37	83	6.2	970	1 US-08-375-709-7	Sequence 7, Appl
38	83	6.2	970	1 US-08-752-929-7	Sequence 7, Appl
39	83	6.2	970	4 US-09-090-793-5	Sequence 5, Appl
40	83	6.2	971	4 US-08-867-611-52	Sequence 52, Appl
41	83	6.2	973	4 US-08-867-611-53	Sequence 53, Appl
42	83	6.2	992	4 US-08-867-611-54	Sequence 54, Appl
43	83	6.2	1786	4 US-08-444-818-54	Sequence 54, Appl
44	83	6.2	2261	4 US-08-444-818-66	Sequence 66, Appl
45	83	6.2	2772	4 US-08-444-818-89	Sequence 89, Appl

ALIGNMENTS

```
RESULT 1
US-09-277-716-22
: Sequence 22, Application US/09277716A
: Patent No. 6232107
: GENERAL INFORMATION:
: APPLICANT: Bryan, Bruce
: APPLICANT: Szent-Gyorgyi, Christopher
: APPLICANT: PROUME, LTD.
: TITLE OR INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
: CURRENT APPLICATION NUMBER: US/092777, 716A
: EARLIER FILING DATE: 1999-03-26
: EARLIER APPLICATION NUMBER: 60/102, 939
: EARLIER FILING DATE: 1998-10-01
: EARLIER APPLICATION NUMBER: 60/089, 367
: EARLIER FILING DATE: 1998-06-15
: EARLIER APPLICATION NUMBER: 60/079, 624
: EARLIER FILING DATE: 1998-03-27
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 22
: LENGTH: 382
: TYPE: PRT
: ORGANISM: Artificial Sequence: fusion protein
: FEATURE:
: OTHER INFORMATION: Cellulose Binding Domain (CBD)-Gausia luciferase fusion prote
US-09-277-716-22

Query Match          36.0% Score 480.5; DB 4; Length 382;
Best Local Similarity 62.6%; Pred. No. 4.1e-41;
Matches 107; Conservative 10; Mismatches 33; Indels 21; Gaps 4;

QY 1 MSVEFYNSNKSQNTSITPTITNTSDSLNLDNVKRYRYTSGTGOFMCDHAGAL 60
    |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1 MSVEFYNSNKSQNTSITPTITNTSDSLNLDNVKRYRYTSGTGOFMCDHAGAL 60
    |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 61 LGNSVDNMTSKYANFVKETASPTSTYDTYLDPSHMRCL--QGSSLIITISVFLVGAQ 117
    |||||||||||||||||||||||||||||||||||||||||||||||
QY 118 NFSRO----EYGAATGAVGCVAAQLGKSGGRVSMATGGA-VLGGIGS 162
    |||||
DB 121 NTQTNDYSFPASSSTFPVNPRTVGT-----IGCAKVLGTAPGS 159
    |||||

RESULT 2
US-08-048-164A-2
: Sequence 2, Application US/08048164A
: Patent No. 5496934
: GENERAL INFORMATION:
: APPLICANT: Shoseyov, Oded
```

APPLICANT: Shpiegl, Itai
APPLICANT: Goldstein, Marc A.
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/048,164A
APPLICATION NUMBER: US/08/048,164A
FILING DATE: 14-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-048-164A-2

Query Match 35.7%; Score 477; DB 1; Length 162;
Best Local Similarity 97.8%; Pred. No. 2.7e-41;
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSVEFYNSKNSAQTNSITPIIKITNTSDSLNLDVKVRYRYTSDGTGQTFCWDHAGAL 60
DB 6 MSVEFYNSKNSAQTNSITPIIKITNTSDSLNLDVKVRYRYTSDGTGQTFCWDHAGAL 65
OY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYL 92
DB 66 LGNSYVDNTSKVTANFVKETASPTSTYDYL 97

RESULT 3
US-08-460-462-2
Sequence 2, Application US/08460462
Patent No. 5670623
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shpiegl, Itai
APPLICANT: Goldstein, Marc A.
APPLICANT: Dol, Roy H.
TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,462
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-460-462-2

Query Match 35.7%; Score 477; DB 1; Length 162;
Best Local Similarity 97.8%; Pred. No. 2.7e-41;
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSVEFYNSKNSAQTNSITPIIKITNTSDSLNLDVKVRYRYTSDGTGQTFCWDHAGAL 60
DB 6 MSVEFYNSKNSAQTNSITPIIKITNTSDSLNLDVKVRYRYTSDGTGQTFCWDHAGAL 65
OY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYL 92
DB 66 LGNSYVDNTSKVTANFVKETASPTSTYDYL 97

RESULT 4
US-08-460-457-2
Sequence 2, Application US/08460457
Patent No. 5719044
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shpiegl, Itai
APPLICANT: Goldstein, Marc A.
APPLICANT: Dol, Roy H.
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN FUSION PROTEINS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,457
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-460-457-2

Query Match 35.7%; Score 477; DB 1; Length 162;
Best Local Similarity 97.8%; Pred. No. 2,7e-41;
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MSVEFYNSKSAQNTSTPIIKITNTSDSLNDVKKVRYTSDGTGOTFCMCDHAGAL 60
Db 6 MSVEFYNSKSAQNTSTPIIKITNTSDSLNDVKKVRYTSDGTGOTFCMCDHAGAL 65
Oy 61 LGNSYDNTSKVTANFVKETASPTSTYDTYLD 92
Db 66 LGNSYDNTSKVTANFVKETASPTSTYDTYVE 97

RESULT 5
US-08-460-458-2
; Sequence 2, Application US/08460458
; Patent No. 5738984
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; TITLE OF INVENTION: KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,458
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-460-458-2

Query Match 35.7%; Score 477; DB 1; Length 162;
Best Local Similarity 97.8%; Pred. No. 2,7e-41;
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MSVEFYNSKSAQNTSTPIIKITNTSDSLNDVKKVRYTSDGTGOTFCMCDHAGAL 60
Db 6 MSVEFYNSKSAQNTSTPIIKITNTSDSLNDVKKVRYTSDGTGOTFCMCDHAGAL 65

Oy 61 LGNSYDNTSKVTANFVKETASPTSTYDTYLD 92
Db 66 LGNSYDNTSKVTANFVKETASPTSTYDTYVE 97

RESULT 6
US-08-460-455-2
; Sequence 2, Application US/08460455
; Patent No. 5837814
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,455
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-460-455-2

Query Match 35.7%; Score 477; DB 2; Length 162;
Best Local Similarity 97.8%; Pred. No. 2,7e-41;
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MSVEFYNSKSAQNTSTPIIKITNTSDSLNDVKKVRYTSDGTGOTFCMCDHAGAL 60
Db 6 MSVEFYNSKSAQNTSTPIIKITNTSDSLNDVKKVRYTSDGTGOTFCMCDHAGAL 65
Oy 61 LGNSYDNTSKVTANFVKETASPTSTYDTYLD 92
Db 66 LGNSYDNTSKVTANFVKETASPTSTYDTYVE 97

RESULT 7
US-08-330-394A-2
; Sequence 2, Application US/08330394A
; Patent No. 5856201
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Yosef, Karney
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.

Tue Oct 29 08:31:12 2002

us-09-677-374-6.oligo.rai

APPLICANT: DOJ, Roy H.
TITLE OF INVENTION: METHODS OF DETECTION USING THE
CELLULOSE BINDING DOMAIN
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/330,394A
FILING DATE: 27-OCT-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CIP OF PCT/US94/04132
FILING DATE: 14-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REFERENCE/DOCKET NUMBER: 7809-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66441 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-330-394A-2

Query Match 35.7%; Score 477; DB 2; Length 162;
Best Local Similarity 97.8%; Pred. No. 2.7e-41;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYSDDGTGQTFWCDHAGAL 60
DB 6 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYSDDGTGQTFWCDHAGAL 65
QY 61 LGNSYVDNNTSKVTANFVKETASPTSTYDTYID 92
DB 66 LGNSYVDNNTSKVTANFVKETASPTSTYDTYVE 97

RESULT 8
US-09-006-636-7
Sequence 7, Application US/09006636
Patent No. 6005092
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
ATTORNEY/AGENT INFORMATION:
NAME: Shani, Ziv
TITLE OF INVENTION: ARABIDOPSIS THALIANA ENDO-1,4-BETA-
GLUCANASE GENE, PROMOTER AND PROTEIN
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/006,636
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7809-019
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 163 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-006-636-7

Query Match 35.7%; Score 477; DB 3; Length 163;
Best Local Similarity 97.8%; Pred. No. 2.7e-41;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYSDDGTGQTFWCDHAGAL 60
DB 7 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYSDDGTGQTFWCDHAGAL 66
QY 61 LGNSYVDNNTSKVTANFVKETASPTSTYDTYID 92
DB 67 LGNSYVDNNTSKVTANFVKETASPTSTYDTYVE 98

RESULT 9
US-09-006-632-7
Sequence 7, Application US/09006632
Patent No. 6184440
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
ATTORNEY/AGENT INFORMATION:
NAME: Shiguel, Etai
TITLE OF INVENTION: TRANSGENIC PLANTS OF ALTERED
MORPHOLOGY
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,632
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7809-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 163 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-006-632-7

Query Match 35.7%; Score 477; DB 4; Length 163;
Best Local Similarity 97.8%; Pred. No. 2.7e-41;
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVEFYNSKSAQNTSITPIIKITNTSDSLNLDVKKVRYTSDGTGOTFCMDHAGAL 60
|||||
DB 7 MSVEFYNSKSAQNTSITPIIKITNTSDSLNLDVKKVRYTSDGTGOTFCMDHAGAL 66

QY 61 GNSYVDNTSKVTANFVKETASPTSTYDTYLD 92
|||||
DB 67 GNSYVDNTSKVTANFVKETASPTSTYDTYVE 98

RESULT 10
US-08-330-394A-29

Sequence 29, Application US/08330394A
Patent No. 5856201

GENERAL INFORMATION:

APPLICANT: Shoseyov, Oded

APPLICANT: Josef, Karmey

APPLICANT: Shpiegl, Itai

APPLICANT: Goldstein, Marc A.

APPLICANT: Dol, Roy H.

TITLE OF INVENTION: METHODS OF DETECTION USING THE

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 AVENUE OF THE AMERICAS

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/330,394A

FILING DATE: 27-OCT-1994

PRIOR APPLICATION DATA: CIP OF PCT/US94/04132

APPLICATION NUMBER: 7809-005

ATTORNEY/AGENT INFORMATION:

NAME: MISROCK, S. LESLIE

REFERENCE/DOCKET NUMBER: 7809-005

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864

TELEX: 66441 PENNIE

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 154 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-330-394A-29

Query Match 35.4%; Score 472; DB 2; Length 154;

Best Local Similarity 97.8%; Pred. No. 8.2e-41;

Matches 89; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 SVEFYNSKSAQNTSITPIIKITNTSDSLNLDVKKVRYTSDGTGOTFCMDHAGAL 61
|||||
DB 1 SVEFYNSKSAQNTSITPIIKITNTSDSLNLDVKKVRYTSDGTGOTFCMDHAGAL 60

QY 62 GNSYVDNTSKVTANFVKETASPTSTYDTYLD 92
|||||

DB 61 GNSYVDNTSKVTANFVKETASPTSTYDTYVE 91

RESULT 11

US-08-330-394A-22

Sequence 22, Application US/08330394A

Patent No. 5856201

GENERAL INFORMATION:

APPLICANT: Shoseyov, Oded

APPLICANT: Josef, Karmey

APPLICANT: Shpiegl, Itai

APPLICANT: Goldstein, Marc A.

APPLICANT: Dol, Roy H.

TITLE OF INVENTION: METHODS OF DETECTION USING THE

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 AVENUE OF THE AMERICAS

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/330,394A

FILING DATE: 27-OCT-1994

PRIOR APPLICATION DATA: CIP OF PCT/US94/04132

APPLICATION NUMBER: 7809-005

ATTORNEY/AGENT INFORMATION:

NAME: MISROCK, S. LESLIE

REFERENCE/DOCKET NUMBER: 7809-005

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864

TELEX: 66441 PENNIE

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 156 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-330-394A-22

Query Match 35.4%; Score 472; DB 2; Length 156;

Best Local Similarity 97.8%; Pred. No. 8.4e-41;

Matches 89; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 SVEFYNSKSAQNTSITPIIKITNTSDSLNLDVKKVRYTSDGTGOTFCMDHAGAL 61
|||||
DB 1 SVEFYNSKSAQNTSITPIIKITNTSDSLNLDVKKVRYTSDGTGOTFCMDHAGAL 60

QY 62 GNSYVDNTSKVTANFVKETASPTSTYDTYLD 92
|||||
DB 61 GNSYVDNTSKVTANFVKETASPTSTYDTYVE 91

RESULT 12

PCT-US95-13813-9

Sequence 9, Application PC/TUS9513813

GENERAL INFORMATION:

APPLICANT: Yeda Research and Development Co. Ltd.

APPLICANT: Ramot University Authority for Applied

APPLICANT: Research and Industrial Development Ltd.

APPLICANT: Bayer, Edward A.

APPLICANT: Moray, Ely

APPLICANT: Wilchek, Meir

[illegible]

```

GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daneels, Roy
Morgan, Hugh W.
Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for
treating Cellulose Containing
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <unknown>
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 1426 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO. 6294366e
SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-09-136-574A-43
Query Match 11.2%; Score 149.5; DB 4; Length 1426;
Best Local Similarity 32.9%; Pred. No. 2.4e-06;
Matches 50; Conservative 16; Mismatches 35; Indels 51; Gaps 6;
OY 1 MSVEFYNSNKSNOTSITPIIKITNTSDSLNDLVKRYVYTSDGTGGFWCDHA--G 58
DB 414 LKVLTKNNETASIGSIRPMFKIVNGSSVDLSRKTRWTTVGDKRQGSACVDMADIG 473
OY 59 ALLGNSTVDNTSKVTANRVKETASTPTSTYYDYRLDPDSHMRGLQGSSLIIISVELGAQN 118
DB 474 A-----SNVFENEVK-LSSCVSGADAYL----- 495
OY 119 FSRQEVGAATGAVVGVAGOLF-GKGSGRVSM 149
DB 496 ---EVSFGSG---AGQLDPGKDAGDIQV 517

```

Search completed: October 27, 2002, 11:12:14
Job time : 14.2642 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:58:51 : Search time 34.0449 Seconds
(without alignments)
835.217 Million cell updates/sec

Title: US-09-677-374-6
Perfect score: 1335
Sequence: 1 MSVERFYNKSAQTNSITPI.....IYGTACPDGRMRYISTEK 256

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1335	100.0	256	22	AA81128
2	832	62.3	161	22	AA81127
3	815	61.0	162	22	AA81126
4	815	61.0	162	22	AA81126
5	487	36.5	190	22	AAE05746
6	487	36.5	328	22	AAE05749
7	487	36.5	341	22	AAE05747
8	483	36.2	162	15	AAE05748
9	483	36.2	163	22	AAE05745
10	483	36.2	428	22	AAE05748
11	480.5	36.0	382	20	AAV39952

12	477	35.7	162	20	AAW90077
13	472	35.4	154	20	AAW90081
14	472	35.4	156	20	AAW90080
15	245	18.4	167	17	AAE95080
16	245	18.4	476	21	AAV54123
17	245	18.4	493	20	AAV28850
18	245	18.4	493	20	AAV43218
19	245	18.4	531	18	AAW15238
20	245	18.4	1853	19	AAW43108
21	149.5	11.2	1426	20	AAV13492
22	148.5	11.1	1616	20	AAV13494
23	148.5	11.1	1751	20	AAV13493
24	127.5	9.6	1352	22	AAE63962
25	126.5	9.5	1350	22	AAE63963
26	122.5	9.2	551	18	AAW18790
27	113.5	8.5	499	14	AAE42122
28	112	8.4	20	22	AAE81130
29	111	8.3	224	22	AAE20105
30	102.5	7.7	230	20	AAV34487
31	102.5	7.7	230	20	AAV34362
32	102	7.6	782	12	AAE15625
33	101.5	7.6	154	11	AAE05799
34	100.5	7.5	700	12	AAE13227
35	98	7.3	922	20	AAV34597
36	98	7.3	922	20	AAE88419
37	98	7.3	922	21	AAE95548
38	95	7.1	309	22	AAE15806
39	95	7.1	2309	22	AAE65232
40	92	6.9	285	20	AAE41688
41	92	6.9	285	21	AAE44244
42	92	6.9	285	22	AAE29025
43	91.5	6.9	285	21	AAE87280
44	91.5	6.9	285	21	AAE73440
45	91.5	6.9	285	22	AAE93288

ALIGNMENTS

RESULT 1	AA81128	standard: Protein: 256 AA.
ID	AA81128	
XX	AA81128:	
AC	AA81128:	
XX		
DT	11-JUL-2001 (first entry)	
XX		
DE	C17E2 Ospea construct with N-terminal fusion partner.	
XX		
KW	poikilothermic fish; piscirickettsia salmonis; rickettsial pathogen;	
KW	vaccine; Ospea; salmonid rickettsial septicaemia; rickettsial disease;	
KW	SRS; 17E2; fusion construct.	
XX		
OS	piscirickettsia salmonis.	
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	Region	1..95
FT	Region	/label= Undefined_N-terminal_fusion_partner
FT	Region	96..256
FT	Region	/label= C17E2_Osps
FT	Region	/note= "Product of Ospea gene optimised for expression in Escherichia coli"
CA2281913-A1.		
17-MAR-2001.		
17-SEP-1999;	99CA-2281913.	
17-SEP-1999;	99CA-2281913.	
(KAYW/) KAY W W.		

C. cellulovorans C
C. cellulovorans C
C. cellulovorans C
Cellulose binding
A mannanase-linker
pectate lyase-link
pectate lyase CBD
Scaffolding protein
C. thermocellum ce
Truncated cellulase
Truncated cellulase
Amino acid sequenc
Amino acid sequenc
Corrected Bacillus
NK-1 cellulase. B
Ospea B-cell epitop
Moraxella catarrha
Porphyromonas ging
Cellulase AE-1. A
PBOP-2 gene prod.
Novel endoglucanase
Chlamydia pneumonia
Chlamydia pneumonia
Novel human diageno
Drosophila melanog
Human PRO284 (UNO2
Human PRO signal pept
Human secreted pro
Human protein HP10

This invention relates to a method for the protection against infection of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia salmonis*. The method comprises administering an immunogenic fragment of *P. salmonis* specific antigen termed OspA, or an immunogenic antigen, particularly poikilothermic fish, against the bacterial pathogen *P. salmonis*. The method is also useful for protecting against salmon *P. salmonis*. The method is also useful for protecting against salmon *P. salmonis*, septicaemia (SRS) and other rickettsial diseases. The pre rickettsial represents the amino acid sequence of C17E2, a *P. salmonis* O sequence represents the amino acid sequence in *Escherichia coli*, fused to an constructed optimised for expression in *Escherichia coli*, fused to an undefined N-terminal fusion partner. The fusion protein is used in a vaccine to create an anti-OspA antibody response.

Query Match	Best Local Similarity	Matches	256; Conservative	100.0%; Score 1335; DB 22; Length 250	0; Mismatches	0; Indels	0; Gaps
QY	1	MSVEFYNSNKSQAQNSITPIIKITNTSDSLNLNDVKVRYYYTSDGTQSQFQWCDHAGAL	60				
Db	1	MSVEFYNSNKSQAQNSITPIIKITNTSDSLNLNDVKVRYYYTSDGTQSQFQWCDHAGAL	60				
QY	61	LGNSIVYDNTSKVTANFYKETAAPTSTVDYFLDPSHMRGCLQGSSLIITISVFLVGCANONS	12				
Db	61	LGNSIVYDNTSKVTANFYKETAAPTSTVDYFLDPSHMRGCLQGSSLIITISVFLVGCANONS	12				
QY	121	ROEVSAATGAVVGCYAGQLFEGKSGSRVSMATGCAVLGGLIGSKTIGOSMDQDDKIKTNOSL	18				
Db	121	ROEVSAATGAVVGCYAGQLFEGKSGSRVSMATGCAVLGGLIGSKTIGOSMDQDDKIKTNOSL	18				
QY	181	EYKACQVTRWRNPDTGNSYSVEPRTYQRTNKQERROQYCFREFQKAMTAGOKELTGT	24				
Db	181	EYKACQVTRWRNPDTGNSYSVEPRTYQRTNKQERROQYCFREFQKAMTAGOKELTGT	24				
QY	241	ACPDGGRMOVISTEK	256				
Db	241	ACPDGGRMOVISTEK	256				

RESULT 2	
AAB81127	
ID	AAB81127 standard; Protein; 161 AA.
XX	
XX	
AC	
XX	AAB81127:
DI	11-JUL-2001 (first entry)
XX	
DE	Optimised OSPA protein 17E2 amino acid sequence.
XX	
KM	Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen
KM	vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease
SRS.	
XX	
OS	Piscirickettsia salmonis.
OS	Synthetic.
XX	
FH	Key
FT	Region
	Location/Qualifiers
	109..128

This invention relates to a method for the production of a poliothermic fish by the bacterial pathogen, *Piscirickettsia salmonis*. The method comprises administering an immunogenic amount of a *P. salmonis* specific antigen termed OspA, or an immunogenic fragment of OspA in the form of a vaccine. The method is used for protecting animals, particularly poliothermic fish, against the bacterial pathogen *P. salmonis*. The method is also useful for protecting against salmonid rickettsial septicæmia (SRS) and other rickettsial diseases. The present invention represents optimised *P. salmonis* OspA protein 17/22. The DNA sequence represents optimised *P. salmonis* OspA protein 17/22. The DNA encoding OspA 17/22 (AA86247) has been optimised for expression in *Escherichia coli*. An OspA protein with an N-terminal fusion partner is used in a vaccine to create an anti-OspA antibody response.

Seq	Sequence	161 AA;	62.3%;	Score 832;	DB 22;	Length 161;
	Query Match		100.0%;	Pred. No. 1.6e-73;		
	Best Local Similarity					
	Matches	161;	Conservative	0;	Mismatches	0;
					Indels	0;
					Gaps	0;
QY	96	MGKCLGGSLIIITVFLVYGCANQNSRQEVGAATGAVVGAGOLFGKSGSRVMAIGAV	155			
Db	1	MGKCLGGSLIIITVFLVGCANQNSRQEVGAATGAVVGAGOLFGKSGSRVMAIGAV	60			
QY	156	LGGLIGSKTIGQSMQDQDKIKLNQSLKRVKAGQVTRRRNPDTGNSVSEPRVRYQRYNKE	215			
Db	61	LGGLISKTKIGQSMQDQDKIKLNQSLKRVKAGQVTRRRNPDTGNSVSEPRVRYQRYNKE	120			
QY	216	RRQCYCRFPOQKAMTAGQKQELTYGACQPDGGRMIVSTEK	256			
Db	121	RRQCYCRFPOQKAMTAGQKQELTYGACQPDGGRMIVSTEK	161			

```

RESULT 3
AAG78025
ID AAG78025 standard; protein; 162 AA.
XX
XX AAG78025;
AC
XX
XX
DT 15-JAN-2002 (first entry)
XX
XX piscirickettsia salmonis polypeptide P10.6.
DE
XX piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
KW septicæmia; SRS; surface antigen; vaccine; antibacterial; fish;
KW ATCC VR-1361.
XX
XX piscirickettsia salmonis.
OS
XX
XX

```

PN WO200168865-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 12-MAR-2001; 2001WO-GB01055.
 XX
 PR 11-MAR-2000; 2000GB-0005838.
 PR 01-JUL-2000; 2000GB-0016080.
 PR 01-JUL-2000; 2000GB-0016082.
 PR 29-JUL-2000; 2000GB-0018599.
 XX
 PA (AQUA-) AQUA HEALTH EURO LTD.
 XX
 PI Sismard N. Brouwers H, Jones S, Griffiths S, Valenzuela P;
 PI Burzio L;
 DR WPI: 2001-639050/73.
 DR N-PSDB: AAH79040.
 XX
 PT New nucleic acids encoding an amino acid sequence homologous to the
 PT surface antigen present on Piscirickettsia salmonis are useful to
 PT protect fish against Piscirickettsiosis -
 XX
 PS Claim 6; Fig 5; 25pp; English.
 XX
 CC The invention relates to nucleic acid sequences and the encoded protein
 CC of a least part of the surface antigen present on Piscirickettsia
 CC salmonis for production of a vaccine with antibacterial activity to
 CC protect fish against P. salmonis which causes piscirickettsiosis, also
 CC known as salmonid rickettsial septicaemia.
 CC
 SQ Sequence 162 AA:

Query Match 61.0%; Score 815; DB 22; Length 162;
 Best Local Similarity 98.8%; Pred. No. 7.6e-72;
 Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 XX
 QY 97 RGCLOGSSLIITISVFLVGCACNFRSEVGAATGAVGVAGOLFCKSGRYSMAIGGAVL 156
 Db 3 RGCLOGSSLIITISVFLVGCACNFRSEVGAATGAVGVAGOLFCKSGRYSMAIGGAVL 62
 QY 157 GGLIGSKIGOSMDQDDIKLNQSLKVKAGOVTRMRNDTGNSTSVSPVRYQRYNKOER 216
 Db 63 GGLIGSKIGOSMDQDDIKLNQSLKVKAGOVTRMRNDTGNSTSVSPVRYQRYNKOER 122
 QY 217 RQYCREFOOKAMITAGOKOEIYGTACPODGRMOWISTEK 256
 Db 123 RQYCREFOOKAMITAGOKOEIYGTACPODGRMOWISTEK 162

RESULT 4
 AAB81126
 ID AAB81126 standard; Protein; 162 AA.
 XX
 AC AAB81126;
 XX
 DT 11-JUL-2001 (first entry)
 XX
 DE Ospa antigen amino acid sequence.
 XX
 KM Polikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
 KM vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
 KM SRS.
 XX
 OS Piscirickettsia salmonis.
 XX
 FH Key Location/Qualifiers
 FT 110..129
 FT Region /label= B_cell_epitope
 PN CA2281913-A1.
 XX
 PD 17-MAR-2001.

XX
 PF 17-SEP-1999; 99CA-2281913.
 XX
 PR 17-SEP-1999; 99CA-2281913.
 XX
 PA (KAYN/) KAY W W.
 PA (BURI/) BURIAN J.
 PA (KUZV/) KUZV M A.
 XX
 PI Kay WM, Burian J, Kuzv MA;
 PI
 DR WPI: 2001-316844/34.
 DR N-PSDB: AAF86246.
 XX
 PT Method for protecting polikilothermic fish against salmonid rickettsial
 PT septicaemia and other rickettsial diseases comprises administering a
 PT vaccine containing the Ospa protein of Piscirickettsia salmonis -
 XX
 PS Example 2; Fig 2B; 35pp; English.
 XX
 CC This invention relates to a method for the protection against infection
 CC of a polikilothermic fish by the bacterial pathogen, Piscirickettsia
 CC salmonis. The method comprises administering an immunogenic fragment of a
 CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of
 CC Ospa in the form of a vaccine. The method is used for protecting animals,
 CC particularly polikilothermic fish, against the bacterial pathogen
 CC P. salmonis. The method is also useful for protecting against salmonid
 CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
 CC sequence represents P. salmonis Ospa protein. An Ospa protein with an
 CC N-terminal fusion partner is used in a vaccine to create an anti-Ospa
 CC antibody response.
 CC
 SQ Sequence 162 AA:

Query Match 61.0%; Score 815; DB 22; Length 162;
 Best Local Similarity 98.8%; Pred. No. 7.6e-72;
 Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 XX
 QY 97 RGCLOGSSLIITISVFLVGCACNFRSEVGAATGAVGVAGOLFCKSGRYSMAIGGAVL 156
 Db 3 RGCLOGSSLIITISVFLVGCACNFRSEVGAATGAVGVAGOLFCKSGRYSMAIGGAVL 62
 QY 157 GGLIGSKIGOSMDQDDIKLNQSLKVKAGOVTRMRNDTGNSTSVSPVRYQRYNKOER 216
 Db 63 GGLIGSKIGOSMDQDDIKLNQSLKVKAGOVTRMRNDTGNSTSVSPVRYQRYNKOER 122
 QY 217 RQYCREFOOKAMITAGOKOEIYGTACPODGRMOWISTEK 256
 Db 123 RQYCREFOOKAMITAGOKOEIYGTACPODGRMOWISTEK 162

RESULT 5
 AAE05746
 ID AAE05746 standard; Protein; 190 AA.
 XX
 AC AAE05746;
 XX
 DT 24-SEP-2001 (first entry)
 XX
 DE Clostridium cellulovorans cellulose binding domain-180 (CBD-180).
 XX
 KM Polysaccharide modification; polysaccharide binding domain; PBD; paper;
 KM yarn; fiber; textile; biological crosslinker; mechanical property;
 KM Clostridium cellulovorans cellulose binding domain; wet strength;
 KM durability; elasticity; CBDc105; cellulose binding protein A; CBP A;
 KM CBD-180.
 XX
 OS Clostridium cellulovorans.
 XX
 PN WO200134091-A2.
 XX
 PD 17-MAY-2001.

```

PF 02-NOV-2000; 2000MO-IL00708.
XX 08-NOV-1999; 99US-0164140.
PR 18-NOV-1999; 99US-0166389.
XX
XX (CBDT-) CBD TECHNOLOGIES LTD.
PA (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
XX Levy I, Nussinovitch A, Shoseyov O;
XX WPI: 2001-457121/49.
XX N-PSDB: AAD11043.
XX
XX Preparation of a polysaccharide containing material having at least one
XX desired structural, chemical, physical, electrical and/or mechanical
XX property -
XX
XX Example 1.2; Fig 1e-1g; 121pp; English.
XX
XX The present invention relates to methods and compositions for cross-
XX linking and/or modifying the properties of polysaccharide materials.
XX The method involves treating the polysaccharide structure with a
XX polysaccharide binding domain (PBD) fusion protein. The method is
XX used to alter the structural, chemical, physical, electrical and
XX mechanical properties of polysaccharide materials such as paper,
XX yarns, fibers and textiles; using biological crosslinking agents.
XX The polysaccharide containing materials have improved mechanical
XX properties such as wet strengths, durability and elasticity. The PBD
XX reagent is applied in the forming stage in fluting paper manufacture
XX which eliminates the recycling step. The use of a biological crosslinker
XX improves the recyclability of paper products. The PBD reagent maintains
XX the fine fibers in a slurry therefore resulting in better recovery of
XX raw materials. The PBD molecules are eluted by strong alkaline conditions
XX which enhances the ability of the alkaline glue used in binding to
XX penetrate paper.
XX The present sequence is Clostridium cellulovorans cellulose binding
XX domain (CBD)-180. CBD-180 contains 180 amino acids of C. cellulovorans
XX CBD (CBDoles) of cellulose binding protein A (CBP A).
XX
XX Sequence 190 AA:
XX
XX Query Match 36.5%; Score 487; DB 22; Length 190;
XX Best Local Similarity 46.8%; Pred. No. 1.4e-39;
XX Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;
XX
XX 1 MSVEFYNSKSAQNSITPIIKITNTSDLNLDVAVRYVYSDGTGQTFWCHAGAL 60
XX 1 MSVEFYNSKSAQNSITPIIKITNTSDLNLDVAVRYVYSDGTGQTFWCHAGAL 60
XX
XX 61 LGNSYVDNTSVYANFVKETASPTSTYDYLDPESHMRCLOGSSLLIISVFLGCAQNF 120
XX 61 LGNSYVDNTSVYANFVKETASPTSTYDYLDPESHMRCLOGSSLLIISVFLGCAQNF 120
XX
XX 121 RQEVGAATGAVGAVGAGQLFGKSGRVSMATGAVLGLIGSKIGSGMDQDKIKLNQSL 180
XX 121 RQEVGAATGAVGAVGAGQLFGKSGRVSMATGAVLGLIGSKIGSGMDQDKIKLNQSL 180
XX
XX 93 -----FGFASRAVL-----KKGQFRTIQ----- 111
XX
XX 181 EKVKAQGVTR-WNPDTGNSYSEVPVTRYQRYNKKOHRQOYCFEFOCKAMIAQOKOELY 238
XX 181 EKVKAQGVTR-WNPDTGNSYSEVPVTRYQRYNKKOHRQOYCFEFOCKAMIAQOKOELY 238
XX
XX 112 -----GRITKSDMSMYTQTDNYSFDASSSTPVNP-----KYTGIGGAK--VL 153
XX 112 -----GRITKSDMSMYTQTDNYSFDASSSTPVNP-----KYTGIGGAK--VL 153
XX
XX 239 GRACPOPD 246
XX 239 GRACPOPD 246
XX
XX 154 GTA-PGPD 160
XX 154 GTA-PGPD 160
XX
XX
XX RESULT 6
XX ID AAE05749 standard; Protein; 328 AA.
XX AC AAE05749;
XX XX
XX DT 24-SEP-2001 (first entry)

```

```

XX Chimeric S peptide-cellulose binding domain-S protein.
XX
XX Polysaccharide modification; polysaccharide binding domain; PBD; paper;
XX yarn; fiber; textile; biological crosslinker; mechanical property;
XX wet strength; durability; elasticity; cellulose binding domain; CBD;
XX chimeric protein; S peptide-cellulose binding domain-S protein;
XX SpB-CBD-Sprot; bovine.
XX
XX Chimeric - Clostridium cellulovorans.
XX
XX Key location/qualifiers
XX Region 30..208
XX Region 226..326
XX Region 327
XX Misc-difference 327
XX /label= "Unknown
XX /note= "Encoded by TAG"
XX Misc-difference 327..328
XX /note= "These residues are absent in the sequence shown
XX as SEQ ID NO: 10 in the sequence listing"
XX
XX WO200134091-A2.
XX
XX 17-MAY-2001.
XX
XX 02-NOV-2000; 2000MO-IL00708.
XX
XX 08-NOV-1999; 99US-0164140.
XX 18-NOV-1999; 99US-0166389.
XX
XX (CBDT-) CBD TECHNOLOGIES LTD.
XX (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
XX Levy I, Nussinovitch A, Shoseyov O;
XX WPI: 2001-457121/49.
XX N-PSDB: AAD11046.
XX
XX Preparation of a polysaccharide containing material having at least one
XX desired structural, chemical, physical, electrical and/or mechanical
XX property -
XX
XX Example 4; Fig 4b-4g; 121pp; English.
XX
XX The present invention relates to methods and compositions for cross-
XX linking and/or modifying the properties of polysaccharide materials.
XX The method involves treating the polysaccharide structure with a
XX polysaccharide binding domain (PBD) fusion protein. The method is
XX used to alter the structural, chemical, physical, electrical and
XX mechanical properties of polysaccharide materials such as paper,
XX yarns, fibers and textiles; using biological crosslinking agents.
XX The polysaccharide containing materials have improved mechanical
XX properties such as wet strengths, durability and elasticity. The PBD
XX reagent is applied in the forming stage in fluting paper manufacture
XX which eliminates the recycling step. The use of a biological crosslinker
XX improves the recyclability of paper products. The PBD reagent maintains
XX the fine fibers in a slurry therefore resulting in better recovery of
XX raw materials. The PBD molecules are eluted by strong alkaline conditions
XX which enhances the ability of the alkaline glue used in binding to
XX penetrate paper.
XX The present sequence is S peptide-cellulose binding domain-S protein
XX (SpB-CBD-Sprot), a fusion protein derived from Clostridium cellulovorans
XX and bovine.
XX Note: This sequence is stated to be the same as that shown as
XX SEQ ID NO: 10 in the sequence listing of the specification. However
XX this sequence has 2 additional residues at its C-terminal.
XX
XX Sequence 328 AA:
XX
XX Query Match 36.5%; Score 487; DB 22; Length 328;

```


Best Local Similarity 46.8%; Pred. No. 2.9e-39;
Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;

QY 1 MSVEFYNSKNSAQTNSTIPIIKITNTSDSLNDVAVRRYRYSYDGTGOTFWCDHAGAL 60
DB 35 MSVEFYNSKNSAQTNSTIPIIKITNTSDSLNDVAVRRYRYSYDGTGOTFWCDHAGAL 94
QY 61 LGSNYVDNTSKVTANFKEETASPTSTYDTLDPSPHMGCLGGSSLLIISVFLVCAQNFNS 120
DB 95 LGSNYVDNTSKVTANFKEETASPTSTYDTLDPSPHMGCLGGSSLLIISVFLVCAQNFNS 126
QY 121 ROEVGAATGAVVGVAGOLFEGKSGRYVSMAGAVLGLGSKIGQSDQDKIKLNQSL 180
DB 127 -----FGFASGRATL-----KKGQFTITQ----- 145
QY 181 EKVKAGQVTR--WRNPPTGSSYSEVPRTYORYNKQERRROOYCFEFOQKAMIAQOKEIY 238
DB 146 -----GRITKSDMSNYQTQTDYSFDASSSTPVVNP-----KVTGYIGGAK--VL 187
QY 239 GTACPOPD 246
DB 188 GTA-PCPD 194

RESULT 7
ID AAE05747 standard; Protein; 341 AA.
AC AAE05747;

DT 24-SEP-2001 (first entry)

DE Clostridium cellulovorans CBD cross linker protein (CCP).

KW Polysaccharide modification; polysaccharide binding domain; PBD; paper;
KM yarn; fiber; textile; biological crosslinker; mechanical property;
KW wet strength; durability; elasticity; cellulose binding domain; CBD;
KM CBD cross linker protein; CCP.

OS Clostridium cellulovorans.

FX Key Location/Qualifiers
FH Misc-difference 340..341

FT /note- "Encoded by CCATAGGAT"

FT Misc-difference 341
FT /note- "This residue is absent in the sequence shown
as SEQ ID NO: 6 in the sequence listing"

PN MO200134091-A2.

PD 17-MAY-2001.

PF 02-NOV-2000; 2000WO-IL00708.

PR 08-NOV-1999; 99US-0164140.

PR 18-NOV-1999; 99US-0166389.

PA (CBDT-) CBD TECHNOLOGIES LTD.
PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.

PI Levy I, Nussinovitch A, Shoseyov O;

DR WPI; 2001-457121/49.

DR N-PSDB; AAD11044.

PT Preparation of a polysaccharide containing material having at least one
PT desired structural, chemical, physical, electrical and/or mechanical
PT property

PS Example 1.2; Fig 2b-2e; 121pp; English.

CC The present invention relates to methods and compositions for cross-
CC linking and/or modifying the properties of polysaccharide materials.

CC The method involves treating the polysaccharide structure with a
CC polysaccharide binding domain (PBD) fusion protein. The method is
CC used to alter the structural, chemical, physical, electrical and
CC mechanical properties of polysaccharide materials such as paper,
CC yarns, fibers and textiles, using biological crosslinking agents.
CC The polysaccharide containing materials have improved mechanical
CC properties such as wet strength, durability and elasticity. The PBD
CC reagent is applied in the forming stage in fluting paper manufacture
CC which eliminates the sizing step. The use of a biological crosslinker
CC improves the recyclability of paper products. The PBD reagent maintains
CC the fine fibers in a slurry therefore resulting in better recovery of
CC raw materials. The PBD molecules are eluted by strong alkaline conditions
CC that enhances the ability of the alkaline glue used in binding to
CC penetrate paper.
CC The present sequence is Clostridium cellulovorans cellulose binding
CC domain (CBD) cross linker protein (CCP).
CC Note: This sequence is stated to be the same as that shown as
CC SEQ ID NO: 6 in the sequence listing of the specification. However
CC this sequence has an additional residue at its C-terminal.
XX

SO Sequence 341 AA;

Query Match 36.5%; Score 487; DB 22; Length 341;

Best Local Similarity 46.8%; Pred. No. 3.1e-39;
Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;

QY 1 MSVEFYNSKNSAQTNSTIPIIKITNTSDSLNDVAVRRYRYSYDGTGOTFWCDHAGAL 60
DB 1 MSVEFYNSKNSAQTNSTIPIIKITNTSDSLNDVAVRRYRYSYDGTGOTFWCDHAGAL 60

QY 61 LGSNYVDNTSKVTANFKEETASPTSTYDTLDPSPHMGCLGGSSLLIISVFLVCAQNFNS 120
DB 61 LGSNYVDNTSKVTANFKEETASPTSTYDTLDPSPHMGCLGGSSLLIISVFLVCAQNFNS 120

QY 121 ROEVGAATGAVVGVAGOLFEGKSGRYVSMAGAVLGLGSKIGQSDQDKIKLNQSL 180
DB 93 -----FGFASGRATL-----KKGQFTITQ----- 111

QY 181 EKVKAGQVTR--WRNPPTGSSYSEVPRTYORYNKQERRROOYCFEFOQKAMIAQOKEIY 238
DB 112 -----GRITKSDMSNYQTQTDYSFDASSSTPVVNP-----KVTGYIGGAK--VL 153

QY 239 GTACPOPD 246
DB 154 GTA-PCPD 160

RESULT 8
ID AAR63634 standard; Protein; 162 AA.

AC AAR63634;

DT 05-JUN-1995 (first entry)

DE Cellulose binding domain.

KW Cellulose binding domain; CBD.

OS Clostridium cellulovorans.

PN MO9424158-A.

PD 27-OCT-1994.

PF 14-APR-1994; 94WO-US04132.

PR 14-APR-1993; 93US-0048164.

PA (REGC) UNIV CALIFORNIA.

PA (YISS) YISSUM RES & DEV CO.

PI Dol RH, Goldstein MA, Shoseyov O, Shpiegl I;

XX DR WPI; 1994-341767/42.
 XX N-PSDB; AAO72917.
 XX Isolated cellulose binding domain and fusion proteins - with
 PT applications, including drug delivery, affinity separations, and
 PT diagnostic techniques
 XX Claim 1; Fig 1; 125pp; English.
 PS
 XX AAO72917/R63634 is a novel isolated cellulose binding domain. It pref.
 CC binds crystalline with a Kd ranging from 1.5 microm - 0.5 microm.
 CC pref. with a Kd of <1.2 microm, esp. <1.0 microm. The protein and
 CC the encoding DNA are claimed, as is a CBD fusion protein (FP)
 CC comprising the CBD and a second protein. The second protein is pref.
 CC Protein A, heparinase, a hormone or an enzyme capable of degrading
 CC an environmental pollutant, or an HSP, HSP antibody, cross-reactive
 CC HSP-related protein or an antigenic portion of this. The CBD and FP
 CC may be used in drug delivery, affinity seps, and diagnostic
 CC techniques. CBD nucleic acid may be obtd. from a variety of cell
 CC sources that produce CBDs that bind with high affinity and in a
 CC reversible manner or that produce CBD encoding mRNA. The preferred
 CC source of CBD encoding nucleic acid is C. cellulovorans.
 CC
 XX Sequence 162 AA:
 SQ
 Query Match 36.2%; Score 483; DB 15; Length 162;
 Best Local Similarity 51.7%; Pred. No. 2.7e-39;
 Matches 106; Conservative 9; Mismatches 14; Indels 76; Gaps 4;
 QY 1 MSVEFYNSKNSAQTNSITPIIKITNTSDSLNDLVKRYRYTSGTGQGFPCDHAGAL 60
 DB 6 MSVEFYNSKNSAQTNSITPIIKITNTSDSLNDLVKRYRYTSGTGQGFPCDHAGAL 65
 QY 61 LGNSYVDNTSKVTANFYKETAASPTSTYDYLDPESHMRGCLQSSLIISVFLVGCANFS 120
 DB 66 LGNSYVDNTSKVTANFYKETAASPTSTYDYLDPESHMRGCLQSSLIISVFLVGCANFS 97
 QY 121 ROEYGAATGAVGVAGQLFGKSGRVSMAIGAVLIGLISKIGQSMQODKIKINOSL 180
 DB 98 -----FGFASGRATL-----KKGQFIIIO----- 116
 QY 181 EKVRAGQVTR--WRNPDTGNSYSVE 203
 DB 117 -----GRITKSDMSNYQTNDYSFD 136
 RESULT 9
 ID AAE05745
 AAEO5745 standard; Protein; 163 AA.
 AC AAE05745;
 XX
 DT 24-SEP-2001 (first entry)
 XX
 DE Clostridium cellulovorans cellulose binding domain (CBDclos).
 XX
 KW Polysaccharide modification; polysaccharide binding domain; PBD; paper;
 KW yarn; fiber; textile; biological crosslinker; mechanical property;
 KW Clostridium cellulovorans cellulose binding domain; wet strength;
 KW durability; elasticity; CBDclos; cellulose binding protein A; CBP A.
 XX
 OS Clostridium cellulovorans.
 XX
 PN WO200134091-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 02-NOV-2000; 2000WO-IL00708.
 XX
 PR 08-NOV-1999; 99US-0164140.
 PR 18-NOV-1999; 99US-0166389.
 XX

PA (CBDT-) CBD TECHNOLOGIES LTD.
 PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
 XX
 XX Levy I, Nussinovitch A, Shoseyov O;
 XX
 XX WPI; 2001-457121/49.
 DR N-PSDB; AAD11042.
 XX
 XX Preparation of a polysaccharide containing material having at least one
 PT desired structural, chemical, physical, electrical and/or mechanical
 PT property
 XX
 XX Example 1.1; Page 111-112; 121pp; English.
 XX
 XX The present invention relates to methods and compositions for cross-
 XX linking and/or modifying the properties of polysaccharide materials.
 XX The method involves treating the polysaccharide structure with a
 XX polysaccharide binding domain (PBD) fusion protein. The method is
 XX used to alter the structural, chemical, physical, electrical and
 XX mechanical properties of polysaccharide materials such as paper,
 XX yarns, fibers and textiles, using biological crosslinking agents.
 XX The polysaccharide containing materials have improved mechanical
 XX properties such as wet strengths, durability and elasticity. The PBD
 XX reagent is applied in the forming stage in fluting paper manufacture
 XX which eliminates the sizing step. The use of a biological crosslinker
 XX improves the recyclability of paper products. The PBD reagent maintains
 XX the fine fibers in a slurry therefore resulting in better recovery of
 XX raw materials. The PBD molecules are eluted by strong alkaline conditions
 XX which enhances the ability of the alkaline glue used in binding to
 XX penetrate paper.
 XX The present sequence is Clostridium cellulovorans cellulose binding
 CC domain (CBDclos) of cellulose binding protein A (CBP A).
 CC
 XX Sequence 163 AA:
 SQ
 Query Match 36.2%; Score 483; DB 22; Length 163;
 Best Local Similarity 51.7%; Pred. No. 2.7e-39;
 Matches 106; Conservative 9; Mismatches 14; Indels 76; Gaps 4;
 QY 1 MSVEFYNSKNSAQTNSITPIIKITNTSDSLNDLVKRYRYTSGTGQGFPCDHAGAL 60
 DB 7 MSVEFYNSKNSAQTNSITPIIKITNTSDSLNDLVKRYRYTSGTGQGFPCDHAGAL 66
 QY 61 LGNSYVDNTSKVTANFYKETAASPTSTYDYLDPESHMRGCLQSSLIISVFLVGCANFS 120
 DB 67 LGNSYVDNTSKVTANFYKETAASPTSTYDYLDPESHMRGCLQSSLIISVFLVGCANFS 98
 QY 121 ROEYGAATGAVGVAGQLFGKSGRVSMAIGAVLIGLISKIGQSMQODKIKINOSL 180
 DB 99 -----FGFASGRATL-----KKGQFIIIO----- 117
 QY 181 EKVRAGQVTR--WRNPDTGNSYSVE 203
 DB 118 -----GRITKSDMSNYQTNDYSFD 137
 RESULT 10
 ID AAE05748
 AAEO5748 standard; Protein; 428 AA.
 AC AAE05748;
 XX
 DT 24-SEP-2001 (first entry)
 XX
 DE Clostridium cellulovorans protein A-cellulose binding domain.
 XX
 KW Polysaccharide modification; polysaccharide binding domain; PBD; paper;
 KW yarn; fiber; textile; biological crosslinker; mechanical property;
 KW wet strength; durability; elasticity; protein A-cellulose binding domain;
 KW ProtA-CBD.
 XX
 OS Clostridium cellulovorans.
 XX

FN	Key	Location/Qualifiers
FT	Region	1..263
FT		/label= Protein_A
FT		/note= "This region is derived from PRIT2T vector"
FT	Region	265..426
FT		/label= CBP_A-protein
FT	Misc-difference	427
FT		/label= Unknown
FT		/note= "Encoded by TAG"
FT	Misc-difference	427..428
FT		/note= "These residues are absent in the sequence shown as SEQ ID NO: 8 in the sequence listing"
XX		
PN	MO200134091-A2.	
XX		
PD	17-MAY-2001.	
XX		
PP	02-NOV-2000; 2000WO-IL00708.	
PR	08-NOV-1999; 99US-0164140.	
PR	18-NOV-1999; 99US-0166389.	
XX		
PA	(CBDT-) CBD TECHNOLOGIES LTD.	
PA	(YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.	
XX		
PI	Levy I, Nussinovitch A, Shoseyov O;	
XX		
DR	WPI: 2001-457121/49.	
DR	N-PSDB: MAD11045.	
XX		
PT	Preparation of a polysaccharide containing material having at least one desired structural, chemical, physical, electrical and/or mechanical property -	
PT		
PS	Example 1.3; Fig 3b-3g; 121pp; English.	
XX		
CC	The present invention relates to methods and compositions for cross-	
CC	-linking and/or modifying the properties of polysaccharide materials.	
CC	The method involves treating the polysaccharide structure with a	
CC	polysaccharide binding domain (PBD) fusion protein. The method is	
CC	used to alter the structural, chemical, physical, electrical and	
CC	mechanical properties of polysaccharide materials such as paper,	
CC	yarns, fibers and textiles, using biological crosslinking agents.	
CC	The polysaccharide containing materials have improved mechanical	
CC	properties such as wet strengths, durability and elasticity. The PBD	
CC	reagent is applied in the forming stage in fluidizing paper manufacture	
CC	which eliminates the sizing step. The use of a biological crosslinker	
CC	improves the recyclability of paper products. The PBD reagent maintains	
CC	the fine fibers in a slurry therefore resulting in better recovery of	
CC	raw materials. The PBD molecules are eluted by strong alkaline conditions	
CC	which enhances the ability of the alkaline glue used in binding to	
CC	penetrate paper.	
CC	The present sequence is Clostridium cellulovorans Protein A-cellulose	
CC	binding domain (ProCA-CBD). This sequence contains a part of CBP A	
CC	protein and a part of protein A derived from pRIT2T vector sequence.	
CC	Note: In the sequence listing of the specification. However	
CC	SEQ ID NO: 8 in the sequence listing of the specification. However	
CC	this sequence has 2 additional residues at its C-terminal.	
XX		
SQ	Sequence 428 AA:	
	Query Match 36.2%; Score 483; DB 22; Length 428;	
	Best Local Similarity 51.7%; Pred. No. 1e-38; Mismatches 14; Indels 76; Gaps 4;	
	Matches 106; Conservative 9;	
Dy	1 MSVEFYNSNKSQTINSTPIIKRTITMSDDLNLNDVKRYYYTSDDGTGOTFWCDHAGAL 60	
Dd	270 MSVEFYNSNKSQTINSTPIIKRTITMSDDLNLNDVKRYYYTSDDGTGOTFWCDHAGAL 329	
Oy	61 LGNSTVDMTSKVTAFNVETASPTSTPYDPLDPHSMMRGCLQSSSLIISVFLPGCAQNRS 120	
Dd	330 LGNSTVDMTSKVTAFNVETASPTSTPYDYVE----- 361	

QY	121	REVEGATCAVGVAGAGOLFEGKSGENVSAIGAVLGLGSLGSGMSMOQDKIKLNQSL	180
Db	362	-----GFASGRATL-----KKGFITIQ-----	380
OY	181	EKKAGQVTR--WRNPDTGNSYSVE	203
Db	381	-----GRITKSDMSNTQINDYSFD	400
RESULT 11			
ID	AAV39952	standard: Protein: 382 AA.	
AC	AAV39952;		
DT	13-DEC-1999	(first entry)	
DE	Gaussia Luciferase fusion protein sequence.		
XX			
XX	Luciferase; green fluorescent protein; GFP; screening assay; diagnosis;		
XX	bioluminescence-generating system; toy; cosmetic; fairy dust; beverage;		
XX	body paint; squirt gun; balloon; slimy play material; soap; toothpaste;		
XX	fusion protein.		
XX			
XX	Gaussia sp.		
XX			
XX	WO9949019-A2.		
XX			
XX	30-SEP-1999.		
XX			
XX	26-MAR-1999;	99WO-US06698.	
XX			
XX	27-MAR-1998;	98US-0079624.	
XX	15-JUN-1998;	98US-0089367.	
XX	01-OCT-1998;	98US-0102939.	
XX			
XX	(PROL-) PROLUME LTD.		
XX	(BRYA/) BRYAN B. J.		
XX			
XX	Bryan BJ, Szent-Gyorgyi C;		
XX			
XX	WPI: 1999-580443/49.		
XX	N-PSDB: AA27550.		
XX			
XX	New isolated Renilla mulleri, Gaussia and Pleuromamma luciferase and		
XX	Renilla and Ptilosarcus green fluorescent protein nucleic acids -		
XX			
XX	Disclosure: Page 222-223; 23pp: English.		
XX			
XX			
XX	This sequence represents a luciferase of the invention. The invention		
XX	relates to Renilla mulleri, Gaussia and Pleuromamma luciferase and		
XX	Renilla and Ptilosarcus green fluorescent protein (GFP) nucleic acids and		
XX	proteins. The luciferases and GFPs can be used in		
XX	bioluminescence-generating systems, assays, screening methods, diagnostic		
XX	method and articles of manufacture. They can be expressed using		
XX	e.g. bacterial, yeast, fungal, plant, insect or animal cells. The Renilla		
XX	mulleri, Gaussia and, Pleuromamma luciferase or Renilla or Ptilosarcus		
XX	GFPs can be used in e.g. toys, cosmetics, fountains, personal care items,		
XX	fairy dust, beverages, soft drinks, foods, textile products, bubbles,		
XX	balloons, personal items, dentifrices, soaps, body paints, bubble bath,		
XX	ink or paper products. In particular, they can be used in e.g. squirt		
XX	guns, pellet guns, finger paints, foot bags, greeting cards, slimy play		
XX	material, clothing, bubble making toys, bath powders, cosmetics, body		
XX	lotions, gels, body powders, body creams, toothpastes, mouthwashes,		
XX	soaps, body paints, bubble bath, inks, wrapping paper, gelatins, icings,		
XX	frostings, greeting cards, beer, wine, champagne, soft drinks, ice cubes,		
XX	ice, dry ice or fountains. The nucleic acids can also be used to produce		
XX	transgenic fish and plants.		
XX			
XX			
XX	Sequence	382 AA:	
Query Match	36.0%;	Score 480.5;	DB 20; Length 382;
Best Local Similarity	62.6%;	Pred No. 1.6e-38;	

Best Local Similarity	97.88;	Pred. NO. 1e-50,	
Matches	90;	Conservative	2;
		Mismatches	0;
		Indels	0;
		Gaps	0

Best Local Similarity	97.8%;	Pred. NO. 3e-30;	
Matches	2;	Mismatches	0;
Conservative	0;	Indels	0;
Gaps	0;		

OY 62 GNSYVDNNTSKVTANFVKETASPTSTYDYL 92
 DB 61 GNSYVDNNTSKVTANFVKETASPTSTYDYL 91

RESULT 14

ID AAW90080 standard; Protein: 156 AA.

AAW90080;

09-MAR-1999 (first entry)

C. cellulovorans CBD-HSP fusion protein.

KW Cellulose binding domain; CBD; Cbpa; fusion protein; diagnosis; HSP;
 KW Immunoblot; heat-shock protein; cross reactive protein; detection;
 KW antigenic fragment; antibody; insulin-dependent diabetes mellitus;
 KW cellulose; chitin; cellulolytic; amorphogenic.

OS Clostridium cellulovorans.
 OS Synthetic.

PN US5856201-A.

PD 05-JAN-1999.

PF 27-OCT-1994; 94US-0330394.

PR 27-OCT-1994; 94US-0330394.

PR 14-APR-1993; 93US-0048164.

PR 14-APR-1994; 94WO-US04132.

PA (REGC) UNIV CALIFORNIA.

PA (YISS) YISSOM RES & DEV CO.

PI Del RH, Goldstein MA, Shoseyov O, Shpiegl I;

DR WPI; 1999-105130/09.

XX Detection of a specific analyte by reaction with binding agent fused

PT to cellulose binding domain - and subsequent treatment with

PT cellulose and reaction of insoluble product with a label specific

PT for the analyte

XX Claim 4; Column 51-52; 63pp; English.

XX This sequence represents a fusion protein composed of the Cbpa protein

CC cellulose binding domain (CBD) from Clostridium cellulovorans and a heat

CC shock protein (HSP) fragment. This protein is used in diagnostic

CC immunoassays, e.g. to detect heat-shock proteins (HSP) and their

CC cross-reactive proteins, antigenic fragments or HSP-specific antibodies

CC to it. The CBD has very high affinity for cellulose (including

CC crystalline forms) and chitin (dissociation constant 0.8-1.4 mu M), but

CC has almost no cellulolytic or amorphogenic activities. The CBD binds over

CC a wide pH range and is not released from cellulose by washing with water.

XX Sequence 156 AA;

Query Match 35.4%; Score 472; DB 20; Length 156;

Best Local Similarity 97.8%; Pred. No. 3e-38; Mismatches 0; Gaps 0;

Matches 89; Conservative 2; Indels 0; Gaps 0;

OY 2 SVEFYNSKSAQNTSIPPIIKITNTSDSLNDLVKRYRYTSDGTGQCFWCHAGALL 61

DB 1 SVEFYNSKSAQNTSIPPIIKITNTSDSLNDLVKRYRYTSDGTGQCFWCHAGALL 60

OY 62 GNSYVDNNTSKVTANFVKETASPTSTYDYL 92

DB 61 GNSYVDNNTSKVTANFVKETASPTSTYDYL 91

RESULT 15
 AAR95080
 ID AAR95080 standard; peptide: 167 AA.

AAR95080;

30-OCT-1996 (first entry)

Cellulose binding domain.

KW Cellulose binding domain; CBD; hapten; moiety; biotin; avidin;

KW streptavidin; affinity chromatography; cell separation,

KW cell immobilisation; protein immobilisation; enzyme immobilisation;

KW multienzyme reactors; signal immunoassays; drug delivery; pesticide;

KW cellulose; chitin.

OS Clostridium thermocellum.

PN WO9613524-A1.

PD 09-MAY-1996.

PF 26-OCT-1995; 95WO-US13813.

PR 27-OCT-1994; 94IL-0111415.

PA (TECR) TECHNION RES & DEV FOUND LTD.

PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.

PA (YEDA) YEDA RES & DEV CO LTD.

PI Bayer EA, Lamed R, Morag E, Shoham Y, Wilchek M;

DR WPI; 1996-239453/24.

XX Modified cellulose-binding domain mols. - having attached hapten,

PT partic. biotin, to provide an affinity system for, e.g. sepsis,

PT assays, reactors, delivery etc.

XX Claim 12; Page 34-35; 53pp; English.

XX A modified cellulose binding domain (CBD) or fraction of it, linked

CC to a hapten moiety via one or more Cys or Lys residues and complexes

CC comprising the biotinylated CBD and a biotin-binding molecule

CC selected from modified or unmodified avidin or streptavidin or an

CC anti-biotin antibody, can be used in affinity chromatography, cell

CC separation, cell, protein and enzyme immobilisation, selective

CC removal of biological materials, multienzyme reactors, signal

CC immunoassays and drug (e.g. pesticide) delivery. The CBD molecule

CC can be modified with the hapten without affecting its high affinity

CC for cellulose, or chitin.

XX Sequence 167 AA;

Query Match 18.4%; Score 245; DB 17; Length 167;

Best Local Similarity 51.5%; Pred. No. 6e-16;

Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;

OY 1 MSVEFYNSKSAQNTSIPPIIKITNTSDSLNDLVKRYRYTSDGTGQCFWCHAGALL 60

DB 9 LKVEFYNSNPSPDTTNSINPORKVNTGSSAIDLSKLTIRYTYVDGQKDFWCHAGALL 67

OY 61 LGN--SYVDNNTSKVTANFVKETASPTSTYDYLPSHMGCLQ 101

DB 68 IGSNGSYVNCITSNWKGFYVKNSS--TNNADRYLEISFGCL 109

Search completed: October 27, 2002, 11:09:12

Job time : 35.0449 secs

THIS PAGE BLANK (USPTO)

Qy	97	RGCLDGLSSLIITISVFLVGCACAFNFSRQEKVGAATGAVGVGAGQLFGKSGGRSMAIGAVL	156
Db	3	RGCLDGLSSLIITISVFLVGCACAFNFSRQEKVGAATGAVGVGAGQLFGKSGGRSMAIGAVL	62
Qy	157	GGLIGSKTIGQSMDDQDKTKLQMSLEKVAAGVTRMRNPDGNSVSVBPVRYQYRNNQER	216
Db	63	GGLIGSKTIGQSMDDQDKTKLQMSLEKVAAGVTRMRNPDGNSVSVBPVRYQYRNNQER	122
Qy	217	ROOYCRREFQOKAMTAGOKOEITYGACPPDGRMVOVISTEK	256
Db	123	ROOYCRREFQOKAMTAGOKOEITYGACPPDGRMVOVISTEK	162


```

DE 17 KDA ANTIGEN (FRAGMENT).
OC male-killing Rickettsia from Adalia decempunctata.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=120393;
RN [1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=20575219; PubMed=11133455;
RA Schuilenburg H.J.G.V.D., Habig M., Sloggett J.J., Webberley M.K.,
RA Bertrand D., Hurst G.D.D., Materus M.E.N.;
RT "Incidence of male-killing Rickettsia spp. (alpha-Proteobacteria) in
RT the Ten-Spot Ladybird Beetle Adalia decempunctata L. (Coleoptera:
RT Coccinellidae).";
RL Appl. Environ. Microbiol. 67:270-277(2001).
DR EMBL; AJ269516; CAB96381.1; -.
FT NON TER 1 1
FT NON TER 144 144
SQ SEQUENCE 144 AA; 14801 MW; C825472F16A56AE7 CRC64;

Query Match 18.8%; Score 251.5; DB 2; Length 144;
Best Local Similarity 39.4%; Pred. No. 3.7e-13;
Matches 52; Conservative 24; Mismatches 45; Indels 11; Gaps 4;

OY 120 SRQEPGATGCAVGVCAQLRFGKSGRSMATG-GAVILGLGKSGSGMQDQR-----I 174
Db 17 NKQGFTLLLGAGGALLSQFEGKRGOL-VGVGVALLGAVIGGIGAGMDEQDRRLAEL 75
OY 175 KLNQSLKVKKQGVVTRMNPPTGNGSYSEVPVETCYRYNKQERROOYCFEFOOKARIAGOK 234
Db 76 TSGRLALEAPSSSNVEMRNPNNGNGVTPNKTYY-----RNSTGYCRHYTOTVYIGGKQ 130
OY 235 QEIVGTACPPDP 246
Db 131 QKSYGNACRPDP 142

RESULT 8
O45996 PRELIMINARY; PRT; 1546 AA.
AC O45996;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SCAFFOLDING PROTEIN PRECURSOR.
GN ClpC.
OS Clostridium cellulolyticum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1521;
RN [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35319;
RX MEDLINE=96218696; PubMed=8636029;
RA Pages S., Belaich A., Tardif C., Reverbel-Leroy C., Gaudin C.,
RA Belaich J.P.;
RT "Interaction between the endoglucanase CelA and the scaffolding
RT protein ClpC of the Clostridium cellulolyticum cellulosome.";
RL J. Bacteriol. 178:2279-2286(1996).
RN [2]
RN RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35319;
RX MEDLINE=99173902; PubMed=10074072;
RA Pages S., Belaich A., Fierobe H.P., Tardif C., Gaudin C.,
RA Belaich J.P.;
RT "Sequence analysis of scaffolding protein ClpC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [3]
RN RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35319;
RA Reverbel-Leroy C., Tardif C., Belaich A., Bernadac A., Gaudin C.,
RA Belaich J.;

```

```

RT      *Molecular study and overexpression of the Clostridium cellulolyticum
RT      celCf cellulase gene in Escherichia coli."
RL      Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC 35319;
RA      Pages S., Belatch A., Reverbel C., Tardif C., Fierobe H.P., Gaudin C.,
RA      Belatch J.P.;
RL      Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL: U40345; AAC28899.2; -.
DR      HSSP: Q06851; INBC.
DR      Interpro: IPR001936; CBD_3.
DR      Interpro: IPR002102; Cohesin.
DR      Interpro: IPR003880; Phosphopant_attach.
DR      Pfam: PF00942; CBD_3; 1.
DR      Pfam: PF00963; Cohesin; 8.
DR      Prodom: PD001947; CBD_3; 1.
DR      Prosite: PS00012; PHOSPHOPANTHEINE; UNKNOWN_1.
KM      Signal.
FT      CHAIN
FT      SIGNAL
FT      SIGNAL
FT      CHAIN
SQ      SEQUENCE 1546 AA; 158748 MW; F8651504EC27809F CRC64;

Query Match
Best Local Similarity 18.7%; Score 249; DB 2; Length 1546;
Matches 72; Conservative 32; Mismatches 81; Indels 28; Gaps 7;

OY      1 MSVEFYNSNKAQNSTPIPIKINTSDSLNDVRYRYTSDTQOGTFMCDHAGAL 60
OY      33 VSVQFNNGSSPAASSNSITARKRYNTSGSPINLADLRLRYTTOADKPLTFMCDHAGYM 92
DB      61 LGNSYDNTSKVTANFVKETASPTSTDTYD---PSIMRGCLGSSLIITSVLVCAQ 117
DB      93 GSGNVIDATSTVTSQSF-KAVSPAYTNADHYLEVALNSDAGSLPAGSGTEIQTFRANDWS 151
OY      118 NPSRQEVGAAGAVGVAGVAGOLFQKGS---GRVSMAGVAVLGVLGSKIGSQMDQOD 173
DB      152 NFDOSNDMSYTAH-----GSYMDMKISAFVGGILAVG--STPGGNNPPQDP 197
OY      174 IKLNSLEKVRAG---QVTRMRNPDTGNSYS 201
DB      198 -TINPTISAKAGSPADTKITLTPNGNTFNGIS 229

RESULT 9
OY      053154 PRELIMINARY; PRT; 154 AA.
AC      053154;
DT      01-NOV-1996 (Tremblrel. 01, Created)
DT      01-NOV-1996 (Tremblrel. 01, last sequence update)
DT      01-NOV-1998 (Tremblrel. 08, last annotation update)
DE      (CLONE PRB F15F 1), 5' END CDS (FRAGMENT).
OS      Rickettsia sp.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Rickettsiaceae; Rickettsiidae; Rickettsia.
OX      NCBI_TaxID=789;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=93084757; PubMed=1452660;
RA      Bald R.W., Lloyd M., Stencos J., Ross B.C., Stewart R.S., Dwyer B.;
RA      "Characterization and comparison of Australian human spotted fever
RT      group rickettsiae.";
RL      J Clin. Microbiol. 30:2896-2902(1992).
DR      EMBL: M9391; AAA73386.1; -.
FT      NON_TER
FT      SEQUENCE 154 AA; 15849 MW; F5C35855EDB439D2 CRC64;

Query Match
Best Local Similarity 18.3%; Score 244; DB 2; Length 154;
Matches 55; Conservative 31; Mismatches 54; Indels 16; Gaps 6;

OY      103 SSIIITSV---FLVGC--AONFSRQEVGAATGAVGVAGOLFQKSGRVSMAG-GAVL 156

```

```

DB      5 SKIMITALATSLQACNGPGMKNKOGTGLLGAGGALLGSQFGKGGOL-VGVGVGALL 63
OY      157 GGLISKTIQSGMDQDK-----IKLNSLEKVRKAGVTFWRNPDTGNSYSVEPVRYORYN 212
DB      64 GAVLGQIGAGDEQDRRLAELTTSORALETAPSGSNVEMRNPNDGNSYGVTFPKTYRST 123
OY      213 KOERROQYCREFOQKAMTAGKQEIYGTACPPDGR 248
DB      124 GQD-----CRVYTVGVIGKQKQKAVGNACQPDQ 154

RESULT 10
OY      09F909 PRELIMINARY; PRT; 151 AA.
AC      09F909;
DT      01-MAR-2001 (Tremblrel. 16, Created)
DT      01-MAR-2001 (Tremblrel. 16, last sequence update)
DT      01-MAR-2001 (Tremblrel. 16, last annotation update)
DE      OUTER MEMBRANE PROTEIN (FRAGMENT).
OS      Rickettsia helvetica.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Rickettsiaceae; Rickettsiidae; Rickettsia.
OX      NCBI_TaxID=35789;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Nilsson K., Pahlson C.;
RT      "Novel peptide diagnostic reagent and kit for detection of
RT      rickettsiosis.";
RL      Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AF181036; AAG09427.1; -.
FT      NON_TER
FT      SEQUENCE 151 AA; 15621 MW; B77407B9C71E4B39 CRC64;

Query Match
Best Local Similarity 17.9%; Score 239; DB 2; Length 151;
Matches 55; Conservative 28; Mismatches 54; Indels 16; Gaps 6;

OY      103 SSIIITSV---FLVGC--AONFSRQEVGAATGAVGVAGOLFQKSGRVSMAG-GAVL 156
DB      5 SKIMITALATSLQACNGPGMKNKOGTGLLGAGGALLGSQFGKGGOL-VGVGVGALL 63
OY      157 GGLISKTIQSGMDQDK-----IKLNSLEKVRKAGVTFWRNPDTGNSYSVEPVRYORYN 212
DB      64 GAVLGQIVAGMDQDRRLAELTTSORALEAPSGSNVEMRNPNDGNSYGVTFPKTY----- 119
OY      213 KOERROQYCREFOQKAMTAGKQEIYGTACPP 245
DB      120 -RNSTGYCREYGTGVVIGKQKQKAVGNACROP 151

RESULT 11
OY      09F001 PRELIMINARY; PRT; 131 AA.
AC      09F001;
DT      01-MAR-2001 (Tremblrel. 16, Created)
DT      01-MAR-2001 (Tremblrel. 16, last sequence update)
DT      01-MAR-2001 (Tremblrel. 16, last annotation update)
DE      17 KDA PROTEIN (FRAGMENT).
OS      Rickettsia sp. California 2.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Rickettsiaceae; Rickettsiidae; Rickettsia.
OX      NCBI_TaxID=147259;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CALIFORNIA 2;
RA      Raoult D.;
RT      "A new spf rickettsia isolated from fleas.";
RL      Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CALIFORNIA 2;
RA      Roux V., Raoult D.;
RL      Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

```

```

33 ISVQFNNGSSPISSSSSIARFKVNTSGSPINLADLKLRYFFIODENKQMTFWCHAGYL 92

```

RESULT 15	
Q977Y4	
ID Q977Y4	PRELIMINARY; PRT; 1483 AA

Query Match	16.7%;	Score 223.5;	DB 16;	Length 1483;
Best Local Similarity	28.4%;	Pred. No. 1.5e-09;		
Matches 57; Conservative	22;	Mismatches 49;	Indels 73;	Gaps 4

Search completed: October 27, 2002, 10:59:56
Job time : 31.7392 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:55:57 ; Search time 9,28497 Seconds
(without alignments)
1067.553 Million cell updates/sec

Title: US-09-677-374-6

Sequence: 1 MSVEFYNSKNSAQTNSITPPL.....ITGTACPPQPGRMQVISTEK 256

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	487	36.5	1848	1 CBPA_CLOCL	P38058 clostridium
2	288	21.6	159	1 17KD_RICPR	P16624 rickettsia
3	286	21.4	159	1 17KD_RICJA	O52762 rickettsia
4	285	21.3	159	1 17KD_RICCN	P05372 rickettsia
5	276.5	20.7	159	1 17KD_RICCY	P22882 rickettsia
6	268	20.1	154	1 17KD_RICAU	P50928 rickettsia
7	259	19.4	154	1 17KD_RICPA	P50930 rickettsia
8	259	19.4	154	1 17KD_RICRH	P50931 rickettsia
9	255	19.2	154	1 17KD_RICMO	P50929 rickettsia
10	255	19.1	154	1 17KD_RICAM	P50927 rickettsia
11	245	18.4	772	1 CTRP_CLOTM	Q01866 clostridium
12	245	18.4	1853	1 CTRP_CLOTM	Q06851 clostridium
13	188.5	14.1	879	1 GUNT_CLOTR	Q02934 clostridium
14	188	14.1	986	1 GUNT_CLOTR	P23659 clostridium
15	162.5	12.2	914	1 GUN2_CLOSR	P50900 clostridium
16	156	11.7	80	1 17KD_RICCA	P29697 rickettsia
17	143.5	10.7	1039	1 GUNB_CALSA	P10474 c endoglucan
18	143.5	10.7	1331	1 MANB_CALSA	P22533 caldocellum
19	138.5	10.4	1742	1 GUNB_CALSA	P22533 caldocellum
20	121	9.1	499	1 GUN2_BACSU	P10475 bacillus su
21	115	8.6	499	1 GUN2_BACSU	P23549 bacillus su
22	112.5	8.4	155	1 PCP_YEREN	P1484 yersinia en
23	111	8.3	145	1 YCEA_PAEUA	P29718 paenibacilli
24	109.5	8.2	499	1 GUN1_BACSU	P07983 bacillus su
25	109	8.2	155	1 SLVB_SALTY	O53549 salmonella
26	104.5	7.8	155	1 SLVB_ECOLI	P55741 escherichia
27	103	7.7	504	1 GUNW_ERMCA	O59395 erwina car
28	101.5	7.6	155	1 PCP_HAETN	P10335 haemophilus
29	101.5	7.6	505	1 GUNW_ERMCA	O47096 erwina car
30	100.5	7.5	700	1 GUNB_PAEUA	P29719 paenibacilli
31	100	7.5	700	1 YCFU_ECOLI	P37716 escherichia
32	100	7.5	444	1 GUNW_ERMCA	O59394 erwina car
33	98	7.3	922	1 PMP1_CHLPH	O92995 chlamydia p

34	97.5	7.3	541	1 NU57_YEAST	P48837 saccharomyc
35	95	7.1	1585	1 YOBQ_BACSU	P45931 bacillus su
36	90	6.7	526	1 KICJ_BOVIN	P06394 bos taurus
37	90	6.7	1161	1 DAN4_YEAST	P47179 saccharomyc
38	90	6.7	1332	1 XKDQ_BACSU	P54334 bacillus su
39	88	6.6	378	1 SLGI_YEAST	P54867 saccharomyc
40	88	6.6	1077	1 HLES_DROME	O02308 drosophila
41	87.5	6.6	1061	1 OAR_MYXXA	P38370 myxococcus
42	86.5	6.5	72	1 OSME_SALTY	P37723 salmonella
43	86.5	6.5	1396	1 VLTF_BPPTS	P13390 bacterioph
44	86.5	6.5	3770	1 ACVS_EPMENI	P27742 emeritella
45	86	6.4	526	1 VP5_BTV11	P33476 bluetongue

ALIGNMENTS

RESULT 1	CBPA_CLOCL	STANDARD:	PRT: 1848 AA.
AC	P38058;		
DT	01-OCT-1994 (Rel. 30, Created)		
DT	01-OCT-1994 (Rel. 30, Last sequence update)		
DT	15-DEC-1998 (Rel. 37, Last annotation update)		
DE	Cellulose binding protein A precursor.		
GN	CBPA.		
OS	Clostridium cellulovorans.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_Taxid=1493;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-9228810; PubMed-1565642;		
RA	Shoseyov O., Takagi M., Goldstein M.A., Doi R.H.;		
RT	"Primary sequence analysis of Clostridium cellulovorans cellulose		
RT	binding protein A.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 89:3483-3487(1992).		
CC	FUNCTION: BINDS TO CELLULOSE FIBERS AND COORDINATES CELLULOSE		
CC	ENZYMES.		
CC	-1- PTM: THE N-TERMINUS IS BLOCKED.		
CC	-1- PTM: CONTAINS ABOUT 10% OF CARBOHYDRATES.		
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION		
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -		
CC	THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS		
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY		
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL		
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/		
CC	OR SEND AN EMAIL TO license@sib-sib.ch).		
DR	EMBL: M73817; AAA23218.1; -.		
DR	PIR: A44140; A44140.		
DR	HSSP: Q06851; INBC.		
DR	Interpro: IPR001956; CBD_3.		
DR	Interpro: IPR002102; Cohesin.		
DR	Pfam: PF00942; CBD_3; 1.		
DR	Pfam: PF00963; Cohesin; 9.		
DR	ProDom: PD001947; CBD_3; 1.		
KW	Cellulose degradation; Cell wall; Repeat; Signal; glycoprotein.		
FT	SIGNAL. 29		
FT	CHAIN. 1 1848		
FT	DOMAIN. 29 189		
FT	SEQUENCE. 1848 AA; 189152 MW; 85FA6CE6F771AF1A CRC64;		
QY	Query Match	36.5%;	Score 487; DB 1; Length 1848;
QY	Best Local Similarity	46.8%;	Pred. No. 6.4e-32;
QY	Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;		
DB	1 MSVEFYNSKNSAQTNSITPPIIKITNTSDSLNLDVYVRYRYYSDDGOTGOTFCMDHAGAL 60		
DB	33 MSVEFYNSKNSAQTNSITPPIIKITNTSDSLNLDVYVRYRYYSDDGOTGOTFCMDHAGAL 92		
QY	61 LGNSYVNTSKSVTANFVKETASPTSTYDTYLDPSHMGCGLGSSLLIISFVLVGCAGNFS 120		

```

|||||
Db 93 LGNSYDNTSKVTANPVKETASTPTDYVE----- 124
QY 121 ROEVGAATGAVGVAGOLFGKSGRVSMAIGAVLIGLIGSKIGOSMDQDKIKLNSL 180
Db 125 -----FGFASGRATL-----KKCGFLTIQ----- 143
QY 181 EKVAGQVTR--WRNPDTGNSVSVEPVRTYQRYNKOERROQYCREFOQKAMTAGQKQETIY 238
Db 144 -----GRITKSDWSNTQTNDYSFSDASSSTPVNP-----KVTGTIGGAK--VL 185

QY 239 GTACPOPD 246
Db 186 GTA-PGPD 192

RESULT 2
17KD_RICPR STANDARD; PRT; 159 AA.
ID 17KD_RICPR
AC P16624;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 17 kDa surface antigen precursor.
GN OMP OR RP833.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MADRID E;
RA MEDLINE=89359171; PubMed=2768201;
RA Anderson B.E., Trianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen
RT gene.";
RL J. Bacteriol. 171:5199-5201(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MADRID E;
RA MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naestlund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M28482; AAA26378.1; ALT_SEQ.
CC EMBL: AJ235273; CAAL5258.1; -.
CC PIR: D33971; D33971.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT SEQUENCE 159 AA; 16672 MW; A33D404B65EB071 CRC64;
SQ

Query Match 21.6%; Score 288; DB 1; Length 159;
Best Local Similarity 37.9%; Pred. No. 7.2e-17;
Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;
QY 103 SLLIISV---FLVGC--AQNFSROEVGATGAVGVAGOLFGKSGRVSMAIG-GAVL 156

```

```

|||||
Db 5 SKMIIALASMLQACNCGSGMKNKQGTLLGAGGALLGSGFGGKGQL-VGVGVGALL 63
QY 157 GGLIGSKIGOSMDQDK---IKLNSLEKAVAGQVTRRNNDPTGNSVSVEPVRTYQRYN 212
Db 64 GAVLGQIGASMDQDPRLLLETSQRALSSAPSGSNIERNRNDNGHGVTPNKTY----- 119
QY 213 KOERROQYCREFOQKAMTAGQKQETIYGTACPOPDGRMVOYIS 253
Db 120 -RNSGQYCREYQTQTVITIGKQOKTYGNACRQPDGMQVYN 159

RESULT 3
17KD_RICUA STANDARD; PRT; 159 AA.
ID 17KD_RICUA
AC Q52764;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 17 kDa surface antigen precursor.
GN OMP.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YH;
RA MEDLINE=95229950; PubMed=7714214;
RA Furuya Y., Katayama T., Yoshida Y., Kaiho I.;
RT "Specific amplification of Rickettsia japonica DNA from clinical
RT specimens by PCR.";
RL J. Clin. Microbiol. 33:487-489(1995).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D16515; BAA03965.1; -.
CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT SEQUENCE 159 AA; 16554 MW; CDDCE7CEBDCDB41 CRC64;
SQ

Query Match 21.4%; Score 286; DB 1; Length 159;
Best Local Similarity 37.9%; Pred. No. 1.1e-16;
Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;
QY 103 SLLIISV---FLVGC--AQNFSROEVGATGAVGVAGOLFGKSGRVSMAIG-GAVL 156
Db 5 SKMIIALATMLQACNCPGSMNKGCTLLGAGGALLGSGFGKGTGQL-VGVGVGALL 63
QY 157 GGLIGSKIGOSMDQDK---IKLNSLEKAVAGQVTRRNNDPTGNSVSVEPVRTYQRYN 212
Db 64 GAVLGQIGASMDQDPRLLLETSQRALSSAPSGSNIERNRNDNGHGVTPNKTY----- 119
QY 213 KOERROQYCREFOQKAMTAGQKQETIYGTACPOPDGRMVOYIS 253
Db 120 -RNSTGQYCREYQTQTVITIGKQOKAYGNACRQPDGMQVYN 159

RESULT 4
17KD_RICCN STANDARD; PRT; 159 AA.
ID 17KD_RICCN
AC P05372;

```



```

RL      Submitted (DEC-1994) to the EMBL/GenBank/DBD databases.
CC      -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC      anchor (Probable).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC      The European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL, U17008; AAA82040.1;
CC      DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC      KW Outer membrane; Lipoprotein; Antigen; Signal.
CC      FT SIGNAL 1 19 BY SIMILARITY.
CC      FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
CC      FT LIPID 20 20 N-ACTYL DIGLYCERIDE (PROBABLE).
CC      FT NON_TER 154 154
CC      SQ SEQUENCE 154 AA; 15897 MW; 5D06F45FD8D5EEC CRC64;
CC
CC      Query Match 19.4%; Score 259; DB 1; Length 154;
CC      Best Local Similarity 36.5%; Pred. No. 1.7e-14;
CC      Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;
CC
CC      Oy 103 SLLIISY---ELVGC--AQNFSEHVEGATGAVGVGAGQLFGKSGKGVMAIT-GAVL 156
CC      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
CC      Db 5 SKIWIATATSLAQNCNPGGMNKQGTGLLGAGGALLTSQGGKKGQL-VGVGALL 63
CC      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
CC      Oy 157 GGLIATIGQSMDDQD---IKLNSLEVKAGQVVRNRPPTGNSYSEPRYORYN 212
CC      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
CC      Db 64 GAVLGQGLGAGAGDEDDRLAEIETSORALETRAPSSSENEMPNNGVGVTPKITY---- 119
CC      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
CC      Oy 213 KOERROQYCRFEQKAMTAGOKOELTYTATCPQPDGR 248
CC      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
CC      Db 120 -RNSTQCYRETYQTVYIGKQKQKAYGNACIQPDGQ 154
CC      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
CC
CC      RESULT 8
CC      17KD_RICRH STANDARD: PRT; 154 AA.
CC      ID ID 17KD_RICRH STANDARD: PRT; 154 AA.
CC      DT 01-OCT-1996 (Rel. 34, Created)
CC      DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC      DT 01-OCT-1996 (Rel. 34, Last annotation update)
CC      DT 01-OCT-1996 (Rel. 34, Last annotation update)
CC      DE 17 kDa surface antigen precursor (Fragment).
CC      GN OMP.
CC      OS Rickettsia rhidicephali.
CC      OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC      OC Rickettsiaceae; Rickettsiae; Rickettsia.
CC      OX NCBI_Taxid-33992;
CC      RN [1]
CC      RA SEQUENCE FROM N.A., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
CC      RA Stothard D.R., Attached to the EMBL/GenBank/DBD databases.
CC      RL Submitted (AUG-1994) to the EMBL/GenBank/DBD databases.
CC      CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC      CC anchor (Probable).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC      The European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL, U11020; AAB07706.1;
CC      DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC      KW Outer membrane; Lipoprotein; Antigen; Signal.
CC      FT SIGNAL 1 19 BY SIMILARITY.
CC      FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
CC      FT LIPID 20 20 N-ACTYL DIGLYCERIDE (PROBABLE).
CC      FT NON_TER 154 154

```


SO SEQUENCE 154 AA: 15895 MW: 0CF85AD5D96DFEFB CRC64:

Query Match 19.4%; Score 259; DB 1; Length 154;

Best Local Similarity 36.5%; Pred. No. 1.7e-14; Mismatches 53; Indels 16; Gaps 6;

Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

OY 103 SSLLITISV---FLVGC--AQNFSROEVGAATGAVGVAGQLFGKSGSRVMAIG-GAVL 156

DB 5 SKMITIALASMLQACNPGCGMKKOGTGLLGAGGALLSGFGKGGOL-VGVGVGALL 63

OY 157 GGLIGSKIGOSMDQDK---IKLNOSLEKVKAGQVTRMNPDTGNSYSVEPVRYORYN 212

DB 64 GAVLGGOIGAGMDEODRRLAELTSQKALETPAGSGSNVEMRNPNGNYGYITPKTKY----- 119

OY 213 KOERROOYCREFOQKAMIAQOKKEITGTACPPDGR 248

DB 120 -RNSTGOYCREYOTVTVIGGKQKQKAYGNACLPDQ 154

Result 9

17KD_RICMO STANDARD: PRT: 154 AA.

AC P50929;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE 17 kDa surface antigen precursor (Fragment).

GN OMP.

OS Rickettsia montana.

OC Bacteria: Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=33991;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-OHIO 83-441;

RA Stochard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;

RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

CC -I- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid

CC anchor (Probable).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL: U11017; AAB07705.1; -

DR PROSITE: PS00013; PROKAR.LIPOPROTEIN; 1.

KW Outer membrane; Lipoprotein; Antigen; Signal.

FT SIGNAL 1 19

FT CHAIN 20 >154

FT LIPID 20 20

FT NON_TER 154 154

FT SEQUENCE 154 AA: 15881 MW: A09C53B8769E31DA CRC64:

Query Match 19.2%; Score 256; DB 1; Length 154;

Best Local Similarity 36.5%; Pred. No. 2.9e-14;

Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

OY 103 SSLLITISV---FLVGC--AQNFSROEVGAATGAVGVAGQLFGKSGSRVMAIG-GAVL 156

DB 5 SKMITIALASMLQACNPGCGMKKOGTGLLGAGGALLSGFGKGGOL-VGVGVGALL 63

OY 157 GGLIGSKIGOSMDQDK---IKLNOSLEKVKAGQVTRMNPDTGNSYSVEPVRYORYN 212

DB 64 GAVLGGOIGAGMDEODRRLAELTSQKALETPAGSGSNVEMRNPNGNYGYITPKTKY----- 119

OY 213 KOERROOYCREFOQKAMIAQOKKEITGTACPPDGR 248

DB 120 -RNSTGOYCREYOTVTVIGGKQKQKAYGNACLPDQ 154

Result 10

17KD_RICAM STANDARD: PRT: 154 AA.

AC P50927;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE 17 kDa surface antigen precursor (Fragment).

GN OMP.

OS Rickettsia amblyomii.

OC Bacteria: Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=33989;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-MO 85-1084;

RA Stochard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;

RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.

CC -I- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid

CC anchor (Probable).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL: U11013; AAB07704.1; -

DR PROSITE: PS00013; PROKAR.LIPOPROTEIN; 1.

KW Outer membrane; Lipoprotein; Antigen; Signal.

FT SIGNAL 1 19

FT CHAIN 20 >154

FT LIPID 20 20

FT NON_TER 154 154

FT SEQUENCE 154 AA: 15879 MW: E4FB6429D943581 CRC64:

Query Match 19.1%; Score 255; DB 1; Length 154;

Best Local Similarity 36.5%; Pred. No. 3.5e-14;

Matches 57; Conservative 29; Mismatches 54; Indels 16; Gaps 6;

OY 103 SSLLITISV---FLVGC--AQNFSROEVGAATGAVGVAGQLFGKSGSRVMAIG-GAVL 156

DB 5 SKMITIALASTLQACNPGCGMKKOGTGLLGAGGALLSGFGKGGOL-VGVGVGALL 63

OY 157 GGLIGSKIGOSMDQDK---IKLNOSLEKVKAGQVTRMNPDTGNSYSVEPVRYORYN 212

DB 64 GAVLGGOIGAGMDEODRRLAELTSQKALETPAGSGSNVEMRNPNGNYGYITPKTKY----- 119

OY 213 KOERROOYCREFOQKAMIAQOKKEITGTACPPDGR 248

DB 120 -RNSTGOYCREYOTVTVIGGKQKQKAYGNACLPDQ 154

Result 11

CIPB_CLOTM STANDARD: PRT: 772 AA.

AC C01866;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein SL/SL)

DE Cellulose integrating protein B (Fragment).

GN CIPB.

OS Clostridium thermocellum.

OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1515;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=YS;

CC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 CC Clostridium.
 CC NCBI_TaxId=1515;
 RN [1]
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 29-40 AND 42-43.
 RP STPAIN-ATCC_27405;
 RX MEDLINE=93302508; PubMed=8316083;
 RX Gengross U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S.,
 RA Demain A.L.;
 RT "Sequencing of a Clostridium thermocellum gene (c1pa) encoding the
 RT cellulosomal SL-protein reveals an unusual degree of internal
 RT homology.";
 RT Mol. Microbiol. 8:325-334(1993).
 RN [2]
 RN SEQUENCE OF 1820-1853 FROM N.A.
 RP MEDLINE=93209931; PubMed=8458832;
 RX Fujino T., Beguin P., Aubert J.-P.;
 RT "Organization of a Clostridium thermocellum gene cluster encoding the
 RT cellulosomal scaffolding protein C1pa and a protein possibly involved
 RT in attachment of the cellulosome to the cell surface."
 J. Bacteriol. 175:1891-1899(1993).
 RN [3]
 RN X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.
 RP MEDLINE=97238934; PubMed=9083107;
 RX Shimon L.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,
 RA Frolow F.;
 RT "A cohesin domain from Clostridium thermocellum: the crystal
 RT structure provides new insights into cellulosome assembly."
 Structure 5:381-390(1997).
 RN [4]
 RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
 RP MEDLINE=97076134; PubMed=88918451;
 RX Torno J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,
 RA Steitz T.A.;
 RT "Crystal structure of a bacterial family-III cellulose-binding
 RT domain: a general mechanism for attachment to cellulose."
 EMBO J. 15:5739-5751(1996).
 RN [5]
 RN X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
 RP MEDLINE=98022914; PubMed=9402065;
 RX Tavares G.A., Beguin P., Alzari P.M.;
 RA "The crystal structure of a type I cohesin domain at 1.7-A
 RT resolution."
 J. Mol. Biol. 273:701-713(1997).
 CC -I- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
 CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
 CC CELLULOLYTIC ENZYMES.
 CC -I- SUBCELLULAR LOCATION: CELL SURFACE.
 CC -I- DOMAIN: CONTAINS 9 COPIES OF A DOMAIN (COHESIN) OF ABOUT 150
 CC RESIDUES THAT BINDS TO THE DOCKERIN DOMAIN BORN BY THE CATALYTIC
 CC COMPONENTS OF THE CELLULOSE.
 CC -I- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
 CC WELL AS IN OTHER C.THERMOCCELLUM CELLULOSE ENZYMES. THIS DOMAIN
 CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT".

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on the way
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

 CC EMBL: L08665; -; NOT_ANNOTATED_CDS.
 DR EMBL: X67506; CAA47840.1; -;
 DR PIR: S36859; S36859;
 DR PDB: 1ANJ; 23-JUL-97.
 DR PDB: 1A0H; 08-JUL-98.
 DR PDB: 1NBC; 26-SEP-97.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR002102; Cohesin.
 DR InterPro: IPR002105; Dockerin_1.
 DR InterPro: IPR002048; EF-hand.

DR Pfam: PF00942; CBD_3; 1.
 DR Pfam: PF00963; Cohesin; 9.
 DR Pfam: PF00404; Dockerin; 1; 2.
 DR ProDom: PD001947; CBD_3; 1.
 DR ProSite: PS00018; EF_HAND; UNKNOWN; 1.
 DR ProSite: PS00448; CLOS_CELLOSOME_RPT; 2.
 DR Cellulose degradation: Cell wall; glycoprotein; Repeat; Signal;
 KW 3D-structure.
 FT SIGNAL 1 28
 FT CHAIN 29 1853
 FT DOMAIN 364 522
 FT DOMAIN 34 1694
 FT REPEAT 29 182
 FT REPEAT 183 322
 FT REPEAT 560 704
 FT REPEAT 724 866
 FT REPEAT 889 1031
 FT REPEAT 1054 1196
 FT REPEAT 1219 1361
 FT REPEAT 1384 1526
 FT REPEAT 1548 1690
 FT DOMAIN 1791 1846
 FT REPEAT 1791 1846
 FT REPEAT 1824 1847
 FT REPEAT 1615 1615
 FT CONFLICT 1615 1615
 FT SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABFC5372 CRC64;

Query Match 18.4%; Score 245; DB 1; Length 1853;
 Best Local Similarity 51.5%; Pred. No. 4; 1e-12;
 Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;

QY 1 MSVEFYNSKSAQTNSTIPPIKIRITNTSDSLNDVKKVRYRYSDTGQGTFCMDHAGAL 60
 Db 369 LKVEFYNSKSDTNTSINPEKVTNTGSSALDLSKLTIRYYTVDGKQDTPFMDHA-AI 427
 QY 61 LGN-STVDNTSKVTANFKETASPTSTYDYLDPSPHMRCLQ 101
 Db 428 IGSNGSYNCITSNVKGFVKMSSS-TNNADTYLEISFTGTL 469

RESULT 13
 GUNZ_CLOTM STANDARD; PRT; 879 AA.
 ID GUNZ_CLOTM
 AC 002934;
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Endoglucanase I precursor (EC 3.2.1.4) (Egi) (Endo-1,4-beta-glucanase)
 GN CELI.
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1515;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.
 RC STRAIN-NCIB 10682;
 RA Hazlewood G.P., Davidson K., Laurie J.I., Huskisson N.S.,
 RA "Gene sequence and properties of Cell, a family E endoglucanase from
 RT Clostridium thermocellum."
 RL J. Gen. Microbiol. 139:307-316(1993).
 CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOTHEROLYSIS OF 1,4-BETA-
 CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL GLUCAN.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- PATHWAY: CELLULOSE DEGRADATION.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSTYL
 CC HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L04735; AAA20892.1; -
 DR PIR: A47704; A47704.
 DR HSP: P26221; 1TF4.
 DR InterPro: IPR001936; CBD_3.
 DR InterPro: IPR001701; Glyco_hydro_9.
 DR Pfam: PF00942; CBD_3; 2.
 DR Pfam: PF00759; Glyco_hydro_9; 1.
 DR ProDom: PD001947; CBD_3; 1.
 DR ProSite: PS00592; GLYCOSTYL_HYDROL_P9_1; 1.
 DR ProSite: PS00698; GLYCOSTYL_HYDROL_P9_2; 1.
 DR Cellulose degradation; Hydrolyase; Glycosidase; signal.
 KW SIGNAL 1 55
 FT CHAIN 56 879
 FT DOMAIN 56 518
 FT ACT_SITE 729 879
 FT ACT_SITE 448 448
 FT ACT_SITE 486 486
 FT ACT_SITE 495 495
 FT SEQUENCE 879 AA; 97796 MW; 35A60069A514A927 CRC64;

Query Match 14.1%; Score 188.5; DB 1; Length 879;
 Best Local Similarity 28.4%; Pred. No. 7; 3e-08;
 Matches 59; Conservative 20; Mismatches 48; Indels 81; Gaps 9;

QY 3 VEFNYSKSAQTNSTIPPIKIRITNTSDSLNDVKKVRYRYSDTGQGTFCMDHAGAL 62
 Db 742 LQYANGAGATSNINPRKIRITNTGSSALDLSKLTIRYYTVDGKQDTPFMDHA-AI 799
 QY 63 NSYVDNTSKVTANFKETASPTSTYDYLDPSPHMRCLQSSLIISVFLGCAQNFPSRQ 122
 Db 800 N-SVNTGNF-NLSPKEGADT-CL- 821
 QY 123 EYCATGAVVGVAGQLFKSGSRVSMAGCAVIGLIGSKIGSMOQDKIKINQSLK 182
 Db 822 EVG-FGSGAG--TLDPGGSV-----EVDIFRSKED 848
 QY 183 VKAQVTRWRNPDTGNSYSEVPRTYR 210
 Db 849 -----WSNYNSNDYSFKQACLROR 868

RESULT 14
 GUNZ_CLOS STANDARD; PRT; 986 AA.
 ID GUNZ_CLOS
 AC P23659;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Endoglucanase 2 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Thermocellulase) (Avicelase I).
 GN CELZ.
 OS Clostridium stercorarium.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1510;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-36 AND 475-486.
 RC STRAIN-NCIB 11745;
 RX MEDLINE=91066838; PubMed=2250652;
 RA Jauris S., Ruecknagel K.P., Schwarz W.H., Kratzsch P.,
 RA Bräuninger K., Staudenbauer W.L.;
 RA "Sequence analysis of the Clostridium stercorarium celz gene encoding
 RT a thermocellulase (Avicelase I): identification of catalytic
 RT and cellulose-binding domains."
 RL Mol. Gen. Genet. 223:258-267(1990).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic

```

RP [CHARACTERIZATION].
RC STRAIN-NCIB 11754;
RA Bronnenmeier K., Ruecknagel K.P., Staudenbauer W.L.;
RT exo-1,4-beta-glucanase (avicelase II) from the cellulolytic thermophile Clostridium stercoararium."
RL Eur. J. Biochem. 200:379-385(1991).
CC -I- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glycosidic linkages in cellulose and cellobiose, releasing cellobiose from the non-reducing ends of the chains.
CC -I- SIMILARITY: BELONGS TO CELLULOSE FAMILY L (FAMILY 48 OF GLYCOSYL HYDROLASESS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outpost at CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z69359; GAA93280.1; -.
DR HSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPRO00556; GlyCo_hydro.48.
DR Pfam: PFO0942; CBD_3; 1.
DR pfam: PF02011; GlyCO_hydro_48; 1.
DR PRINTS: PR00844; GLHYDLRASEA8.
DR PRODOM: PD001947; CBD_3; 1.
DR ProDom: PD011903; GlyCO_hydro_48; 1.
KW Cellulose degradation; Hydrolases; Glycosidase; Signal.
FT SIGNAL 1 33
FT CHAIN 34 914 POTENTIAL.. EXOGLUCANASE II.
SQ SEQUENCE 914 AA: 103020 MW; DDDB60IYD6DFE82C CRC64;

Query Match          12.2%; Score 162.5; DB 1; Length 914;
Best Local Similarity 37.8%; Prec. No. 1e+05;
Matches   34; Conservative    17; Mismatches     30; Indels      9; Gaps       2

QY        3 VEFNSNRKSAQTINSITPIKITNTSDSLNDVKRYVTSISGTOGGOFWCDHAGALLG 62
           ::::|::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db         IQSNAMNQELSIIMPRFRIVSGNLSIPLESKLYTYTVDGDKPQNFCWMASL--- 825

QY        63 NSYVDNSTKYAVANKETASTSYDYLD 92
           :||| || | : | | | |
Db         826 ----GSSENVTGFVKMDGATTGA-DYLE 849


Search completed: October 27, 2002, 10:56:33
Job time : 12.285 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:56:01 : Search time 16.8014 Seconds

(without alignments)
1464.098 Million cell updates/sec

Title: US-09-677-374-6

Perfect score: 1335
Sequence: 1 MSVEFYNSKNSAQTNSITPPI.....IYGTACPPDPGRNQVISTEK 256

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database :

1: PIR-71:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	487	36.5	1848	2 A44140	cellulose-binding
2	288	21.6	159	2 D33971	ricketsial common
3	285	21.3	159	2 B33971	ricketsial common
4	285	21.3	159	2 A33971	ricketsial common
5	285	21.3	159	2 G97860	17K surface antigen
6	276.5	20.7	159	2 C33971	ricketsial common
7	276	20.7	159	2 A25872	17K antigen precursor
8	249	18.7	586	2 PC6006	scaffolding protein
9	245	18.4	1854	2 S36859	clpA protein - C10
10	233.5	17.5	1162	2 T30433	scaffolding protein
11	223.5	16.7	1483	2 C97012	probably cellulose
12	188.5	14.1	879	2 A47704	endoglucanase I (E
13	188	14.1	986	2 S12021	thermoactive cellu
14	151.5	11.3	1779	2 T31085	xylinase - Caldice
15	147	11.0	915	2 A43802	cellulase (EC 3.2.
16	143.5	10.7	1039	2 S02711	cellulase (EC 3.2.
17	143.5	10.7	1331	2 A48954	mannan endo-1,4-be
18	142	10.6	1711	2 T31337	1,4-beta-glucanase
19	138.5	10.4	1742	2 T17120	cellulase (EC 3.2.
20	127.5	9.6	182	2 D83169	conserved hypochet
21	122	9.1	257	2 F87413	hypochetrical prote
22	121	9.1	508	2 G69593	cellulase (EC 3.2.
23	120.5	9.0	131	2 A13418	17K surface antigen
24	115	8.6	489	2 JN0111	cellulase (EC 3.2.
25	113.5	8.5	499	2 A27198	cellulase (EC 3.2.
26	112.5	8.4	155	2 S23787	outer membrane lip
27	111.5	8.4	154	2 B83514	conserved hypochet
28	111	8.3	145	2 A41897	cellulase homolog
29	109.5	8.2	508	2 A26874	cellulase (EC 3.2.

30	109	8.2	155	2 A10693	outer membrane lip
31	109	8.2	179	2 AE0644	probable secreted
32	109	8.2	232	2 G87629	hypochetrical prote
33	109	8.2	257	2 B82837	conserved hypochet
34	107.5	8.1	142	2 AD2696	lipA protein (lipo
35	107.5	8.1	155	2 AF0289	probable lipoprote
36	107	8.0	125	2 D97478	lipA protein (lipo
37	107	8.0	486	2 I40548	bifunctional cellu
38	106.5	8.0	155	2 AG0443	outer membrane lip
39	104.5	7.8	155	2 C64921	outer membrane lip
40	104.5	7.8	155	2 F90922	probable outer mem
41	104.5	7.8	139	2 B85771	probable outer mem
42	104	7.8	135	2 S58234	lipA protein - Rhl
43	103	7.7	232	2 AD3350	outer membrane pro
44	103	7.7	504	2 S54744	cellulase (EC 3.2.
45	101.5	7.6	155	2 T64130	PAI cross-reacting

ALIGNMENTS

RESULT 1
A44140
cellulose-binding protein A precursor - Clostridium cellulovorans
C:Species: Clostridium cellulovorans
C>Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 15-Oct-1999
C:Accession: A44140
R:Shoseyov, O.; Takagi, M.; Goldstein, M.A.; Dei, R.H.
Proc. Natl. Acad. Sci. U.S.A. 89, 3483-3487, 1992
A:Title: Primary sequence analysis of Clostridium cellulovorans cellulose binding pro
A:Reference number: A44140; MUID:92228810
A:Accession: A44140
A>Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-1848 <SHO>
A:Cross-references: GB:M73817; NID:g144748; PIDN:AAA23218.1; PID:g144749
C:Genetics:
A:Gene: cbpa

Query Match 36.5%; Score 487; DB 2; Length 1848;
Best Local Similarity 46.8%; Pred. No. 2.8e-32;
Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;

QY	1	MSVEFYNSKNSAQTNSITPPIIKITNTSDSLNDLVKRYRTSDGQGFPCWDHAGAL	60
DB	33	MSVEFYNSKNSAQTNSITPPIIKITNTSDSLNDLVKRYRTSDGQGFPCWDHAGAL	92
QY	61	LGNSYVNTSKVTANFYKETASPTSTYDYLDPSPHMRGLQSSLIITSVFLVCAQNF	120
DB	93	LGNSYVNTSKVTANFYKETASPTSTYDYLDPSPHMRGLQSSLIITSVFLVCAQNF	124
QY	121	RQEVGAATGAVGVAGQLFGKSGRVSMAIGAVLGLGSLIGQSDQDQIKLNQSL	180
DB	125	RGFASGRATL-----KKGFITID-----	143
QY	181	EKVAGQVTR--WRNPDTGNSYSEPVRTYRYNKKERQOYCEFOQKAMIAQOKETI	238
DB	144	GRITSDMSNYQTQNDYSPDASSTPVNP-----KVYGIYGA--VL	185
QY	239	GTACPOP 246	
DB	186	GTA-PGPD 192	

RESULT 2
D33971
ricketsial common antigen precursor - Rickettsia prowazekii
N:Alternate names: 17kD surface antigen; outer membrane protein (omp); RP833
C:Species: Rickettsia prowazekii
C>Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 03-Nov-2000
C:Accession: D33971; B71645
R:Anderson, B.E.; Tzianabos, T.
J. Bacteriol. 171, 5199-5201, 1989

RESULT 3
B33971
Rickettsial common antigen precursor - Rickettsia conorii
C:Species: Rickettsia conorii
C:Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 20-Aug-1999
C:Accession: B33971
R:Anderson, B.E.; Tzianabos, T.
J: Bacteriol. 171, 5199-5201, 1989
A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene
A:Reference number: A33971; MUID:89359171
A:Accession: B33971
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-159 <AND>
A:Cross-references: GB:M28479; NID:g152463; PIDN:AAA26379.1; PID:g152464
A:Superfamily: rickettsial common antigen

[illegible]

RESULT 4
A33971
Rickettsial common antigen precursor - Rickettsia rickettsii
C:Species: Rickettsia rickettsii
C:Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 20-Aug-1999
C:Accession: A33971
R:Anderson, B.E.; Tzianabos, T.
J. Bacteriol. 171, 5199-5201, 1989
A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene
A:Reference number: A33971; MUID:89359171
A:Accession: A33971
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-159 <ND>
A:Cross-references: NID:g152457; PIDN:AAA26376.1; PID:g152458
C:Superfamily: rickettsial common antigen

[illegible]

```

RESULT 5
G97860
17k surface antigen precursor [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: G97860
C:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.;
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; M01D:2142074; PMID:1157893
A:Accession: G97860
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-159 <K0U>
A:Cross-References: GB:AE006914; PIDN:AAL03825.1; PTD:915620425; GSPDB:GN00173
C:Genetics:
A:Gene: omp
C:Superfamily: rickettsial common antigen

Query Match      21.3%  Score 285;  DB 2;  Length 159;
Best Local Similarity 37.9%  Pred No. 1, 1e-16;
Matches 61;  Conservative 31;  Mismatches 53;  Indels 16;  Gaps 6;

```

[illegible]

Rickettsial common antigen precursor - Rickettsia typhi

C:Species: Rickettsia typhi

C>Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 20-Aug-1999

C:Accession: C33971

R:Anderson, B.E.; Tzianabos, T.

J. Bacteriol. 171, 5199-5201, 1989

A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.

A:Reference number: A33971; MUID:89359171

A:Accession: C33971

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-159 <AND>

A:Cross-references: GB:M28481; NID:g152459; PIDN:AAA26377.1; PID:g152460

C:Superfamily: rickettsial common antigen

Query Match

20.7%; Score 276.5; DB 2; Length 159;

Best Local Similarity 39.6%; Pred. No. 5.8e-16;

Matches 55; Conservative 27; Mismatches 46; Indels 11; Gaps 4;

OY 120 SRQEVGAATGAVVGVAGOLFQKSGRVSMAIG-GAVLGGLIGSKIGSMDOOK-----I 174

DB 27 NKQGTGLLGAGGALGSGFCHGKGL-VGVGALLGAVLGQICASLDQDRLLEL 85

OY 175 KLNQSLKVKAGQVTRMRNPDGNSVSEPVRYQRYNKOERQOYCRFQOKAMIAQOK 234

DB 86 TSQRALESAPSGSNIEMRNDNCHGVTPNKTY-----RNSTQCYCRETYQTGVVIGKQ 140

OY 235 QEITGTACPPDGRMVOIS 253

DB 141 QTTYGNAKROPDQGMQVYN 159

RESULT 7

A25972 17K antigen precursor - Rickettsia rickettsii

C:Species: Rickettsia rickettsii

C>Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 20-Aug-1999

C:Accession: A25972

R:Anderson, B.E.; Regnery, R.L.; Carlone, G.M.; Tzianabos, T.; McBride, J.E.; Fu, Z.Y.; F

J. Bacteriol. 169, 2385-2390, 1987

A:Title: Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia rickettsii.

A:Reference number: A25972; MUID:87222152

A:Accession: A25972

A:Molecule type: DNA

A:Residues: 1-159 <AND>

A:Cross-references: GB:M16486; NID:g152467; PIDN:AAA26381.1; PID:g152468

C:Superfamily: rickettsial common antigen

Query Match

20.7%; Score 276; DB 2; Length 159;

Best Local Similarity 37.3%; Pred. No. 6.4e-16;

Matches 60; Conservative 31; Mismatches 54; Indels 16; Gaps 6;

OY 103 SLLIISV---FLVGC--AQNFSHOEVAATGAVVGVAGOLFQKSGRVSMAIG-GAVL 156

DB 5 SKIMIALATSMIACNGPGMKNQGTGTLIGAGALLGSGFGKGLQ-VGVGALL 63

OY 157 GGLIGSIQSMDOOK-----IKLNQSLKVKAGQVTRMRNPDGNSVSEPVRYQRYN 212

DB 64 GAVLGQIGAGMDQDRRLAELTSQLALETAPSGSNVEMRNPDGNGYVTPNKTY----- 119

OY 213 KOERROOYCRFQOKAMIAQOKEIYGNACPPDGRMVOIS 253

DB 120 -RNSTQCYCRETYQTGVVIGKQOKAYGDACQPDQGMQVYN 159

RESULT 8

PC6006 scaffolding protein ClpC precursor - Clostridium cellulolyticum (fragment)

C:Species: Clostridium cellulolyticum

C>Date: 17-Aug-1990 #sequence_revision 16-Aug-1996 #text_change 13-Nov-1998

C:Accession: PC6006

R:Pages, S.; Belach, A.; Tardif, C.; Reverbel-Leroy, C.; Gaudin, C.; Belach, J.P.

J. Bacteriol. 178, 2279-2286, 1996

A:Title: Interaction between the endoglucanase CelA and the scaffolding protein ClpC

A:Reference number: PC6006; MUID:96218696

A:Accession: PC6006

A:Molecule type: DNA

A:Residues: 1-586 <PAC>

A:Cross-references: GB:U40345; NID:g1314291; PID:g1314292

A:Experimental source: ATCC 35319

C:Genetics:

A:Gene: clpC

F:1-27/Domain: signal sequence #status predicted <SIG>

F:28-586/Product: scaffolding protein ClpC #status predicted <MAT>

F:28-187/Domain: cellulose binding #status predicted <CBD>

F:188-292/Domain: hydrophilic <HYL>

F:293-439/Domain: hydrophobic <HYB1>

F:440-586/Domain: hydrophobic <HYB2>

Query Match

18.7%; Score 249; DB 2; Length 586;

Best Local Similarity 33.8%; Pred. No. 5.7e-13;

Matches 72; Conservative 32; Mismatches 81; Indels 28; Gaps 7;

OY 1 MSVEFYNSKSAQNSITPIKITNTSDSLNDVYRYYTSDTQGGTFMCDHAGAL 60

DB 33 VSVQFVNGSSPSSNSIYAREKYNTSGSPINLADLRYYTYDDADKPLTFMCDHAGYM 92

OY 61 LGNSVYDNTSKVTANFYKETASPTSTYDTYLD--PSHMRCCLOGSSLIISVFLVCAQ 117

DB 93 SGNVYIDATSKVYTSF-KANSPAVTNADHYLEVALNDAGSLPAGSGEIQTRARNDMS 151

OY 118 NFSRQEVGAATGAVVGVAGOLFQKSGS---GRVSMAIGAVLGGLIGSKIGSMDOOK 173

DB 152 NFSQSDMSVYTA-----GSYMDQKISAFVGTGLAVG--STPDGGRPPDOP 197

OY 174 IKLNQSLKVKAG-----QVTRMRNPDGNSYS 201

DB 198 -TIMPTSIKAGSPADTKITLPNGNTFNGIS 229

RESULT 9

S36859 ClpA protein - Clostridium thermocellum

N:Alternate names: probable cellulosome protein large chain SL

C:Species: Clostridium thermocellum

C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000

C:Accession: S36859; S33527; S25767; S28659; T18261

R:Gerngross, U.T.; Demain, A.L.

Submitted to the EMBL Data Library, January 1993

A:Reference number: S36859

A:Accession: S36859

A:Molecule type: DNA

A:Residues: 1-1854 <GER>

A:Cross-references: EMBL:L08665

Mol. Microbiol. 8, 325-334, 1993

A:Title: Sequencing of a Clostridium thermocellum gene (clpA) encoding the cellulosome

A:Reference number: S33527; MUID:93302508

A:Accession: S33527

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1615, 1617-1854 <GE2>

A:Cross-references: EMBL:L08665

R:Fullino, T.; Beguin, P.; Aubert, J.P.

FEMS Microbiol. Lett. 94, 165-170, 1992

A:Title: Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides tha

A:Reference number: S25767

A:Accession: S25767

A:Molecule type: DNA

A:Residues: 1216-1237, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615, 1617-1854 <FNU>

A:Cross-references: EMBL:X67406

R:Beugin, P.

Submitted to the EMBL Data Library, August 1992

A:Reference number: S28659

A:Accession: S28659

[illegible]

		14.1%;	Score 188.5;	DB 2;	Length 879;	
Query Match		28.4%;	Pred. No. 1;e-07;			
Best Local Similarity						
Matches 59;	Conservative	20;	Mismatches 48;	Indels 81;	Gaps 9;	
OY	3	VEEFSNKSASQTSITPIIRIKITNTSDSLINDLVAVRYRTYSDDGQTFMCHAGALLG	62			
	..:	..:	..:	..:	..:	..:
	..:	..:	..:	..:	..:	..:
Db	742	LYQVANGAGTATSIISPIRKIIINNGKATLISDAVIRIYRTYKEGGAQSNFQCDWSSA	-G	799		
OY	63	NSVYDNTSKYTANFVKEETASPTSYDYTLDPHSRMCLOGSSLIIISVFLGCAQNTSRQ	122			
Db	800	N-----SNTGNFF-NLSPKRGADT-----CL-----	821			
OY	123	EVGAATCAVVGAGVAGGLFCGKSGRVMATGAGVAGLGLISKIGQSMDOQDKIKLNQSLER	182			
Db	822	EVG-----FGSGAG--TLDGGSV-----EVOIRFESKED	848			
OY	183	VKAGQVTRWRNPDTGNSVVEPYRTYOR	210			

Db 849 -----WSNYNOSNDYSEKQACLROR 868

RESULT 13

thermoactive cellulase - Clostridium stercorarium
S12021
N:Alternate names: avicelase I
C:Species: Clostridium stercorarium
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
C:Accession: S12021
R:Jauris, S.; Ruecknagel, K.P.; Schwarz, W.H.; Kratzsch, P.; Bronnenmeier, K.; Staendenbe
Mol. Gen. Genet. 223, 258-267, 1990
A:Title: Sequence analysis of the Clostridium stercorarium celz gene encoding a thermoac
A:Reference number: S12021; MUID:1066838
A:Accession: S12021
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-986 <JAU>

Query Match 14.1% Score 188; DB 2; Length 986;
Best Local Similarity 27.6% Pred. No. 1.4e-07;
Matches 47; Conservative 23; Mismatches 58; Indels 36; Gaps 6;

QY 1 MSVEFYNSKSAQNSTIPPIIKITNTSDSLNDVKRYRYTSDGTGQTFMCDHAGL 60
DB 840 IQIMFNGNTSDKNGIMPRRLNTGTPIRLSDVKIRYYTIDGKDNFMCDMSV- 898
QY 61 LGNSYVDNTSKYVTFANFVKETASPTSTYDTPDPSHMRGC--LQGSLLIIVFLVGAON 118
DB 899 -----GNNITGTFTVK-MAEPKEGADYLETGFTDAGYLOPNOSIEVO-----NR 943
QY 119 FSR-----GEVGAATGAVVGAGQLFGKSGRSMAGAVLGL 159
DB 944 FSKADMTDYIOTNDTSFTNTSYG-----SNDRIYVIGVLVSGI 984

RESULT 14

T31085

xylanase - Caldiceilus disruptor sp.

C:Species: Caldiceilus disruptor sp.

C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

C:Accession: T31085

R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.

submitted to the EMBL data library, December 1997

A:Description: Family 10 and 11 xylanase genes from Caldiceilus disruptor sp. Rt69B.1.

A:Reference number: Z20972

A:Accession: T31085

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1779 <MOR>

A:Cross-references: EMBL:AF036924; NID:g2760905; PID:g2760909; PIDN:AAB95326.1

C:Genetics:

A:Note: xync

Query Match 11.3% Score 151.5; DB 2; Length 1779;
Best Local Similarity 33.6% Pred. No. 0.00032;
Matches 51; Conservative 15; Mismatches 35; Indels 51; Gaps 6;

QY 1 MSVEFYNSKSAQNSTIPPIIKITNTSDSLNDVKRYRYTSDGTGQTFMCDHA--G 58
DB 1114 LKVLKNNETASSTGSIKPKRYVNGSSVDLSVKIRIYTYVDGDKPQSAVCDMAOIG 1173
QY 59 ALLGNSYVDNTSKYVTFANFVKETASPTSTYDTPDPSHMRGCLQGSLLIIVFLVGAON 118
DB 1174 A-----SNVTFNFVKLT-SGVSAGADYLL----- 1195
QY 119 FSRQEVGAATGAVVGAGQLF-GKSGRSM 149
DB 1196 -----EVGFSSG-----AGQLQPKDGTIOV 1217

RESULT 15

A43802

cellulase (EC 3.2.1.4) / cellulose 1,4-beta-D-glucosidase (EC 3.2.1.91) - Caldiceilus

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Caldiceilus saccharolyticum

C:Date: 30-Jan-1993 #sequence_revision 30-Sep-1993 #text_change 10-Jul-1998

C:Accession: A43802

R:Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P

Appl. Environ. Microbiol. 56, 3117-3124, 1990

A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophil

A:Reference number: A43802; MUID:91136262

A:Accession: A43802

A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual

A:Molecule type: DNA

A:Residues: 1-915 <SAU>

A:Cross-references: EMBL:X13602

C:Function: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 11.0% Score 147; DB 2; Length 915;
Best Local Similarity 23.2% Pred. No. 0.00034;
Matches 60; Conservative 35; Mismatches 64; Indels 100; Gaps 11;

QY 1 MSVEFYNSKSAQNSTIPPIIKITNTSDSLNDVKRYRYTSDGTGQTFMCDHA--G 58
DB 351 IKVLYANKETNTSTNTTRPWLKRVNAGSSSIDLSKRTIRIYTYVDGERAQSAVSUMAIG 410
QY 59 ALLGNSYVDNTSKYVTFANFVKETASPTSTYDTPDPSHMRGCLQGSLLIIVFLVGAON 118
DB 411 A-----SNVTFNFVK-LSSVSGADYLL----- 432
QY 119 FSRQEVGAATGAVVGAGQLF-GKSGRSMAGAVLGLGSLGSGQSDQDOKILN 177
DB 433 -----EIGFKSG-----AGQLQPKDGTGEI-----QIRFN 457
QY 178 QSLKRVKAGQYTRMRNPTDGTGNSYSEVRYTQRYNKKOEROOYCFEPQKAMINGOKEI 237
DB 458 KS-----DMSNINQGDMSW--LQSMSTYGENEKVTAYT---DGLVWQGE--- 498
QY 238 YGTACPPQPDGRWQVISTEK 256
DB 499 -----PSTNDDMIVYSGNK 512

Search completed: October 27, 2002, 10:58:43
Job time : 19.8014 secs

THIS PAGE BLANK (USPTO)